

# PS 1 QTM 200

Robert Song

1/27/2020

```
#Setting work directory
setwd("~/GitHub/QTM200Spring2020/problem_sets/PS1/My problem set")

#Problem 1:

#Taking Data set and pasting
problem_1 <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113,
112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)

mean(problem_1) #the sample mean for IQ score is 98.44

## [1] 98.44

length(problem_1) # there are 25 observations by counselor on student IQ
scores

## [1] 25

sd(problem_1) # the standard deviation of the observations in IQ scores is
13.09

## [1] 13.09287

std_error <- sd(problem_1) / sqrt(length(problem_1)) #accounting standard
deviation based on our sample size to obtain sample error
std_error # standard error is 2.62

## [1] 2.618575

t.test(problem_1, alternative = "two.sided", conf.level= 0.90) #Running a t
test because we have less than 30 observations, with a desired CI of 90
percent

##
## One Sample t-test
##
## data:  problem_1
## t = 37.593, df = 24, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  93.95993 102.92007
## sample estimates:
## mean of x
##      98.44
```

*#the average student IQ scores is between 93.96 and 102.92 90 percent of the time*  
*#in other words, we are 90 percent confident that the average student IQ score would be between 93.96 and 102.92*

## #Problem 2

```
problem_2 <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)
```

```
t.test(problem_2, mu = 100, alternative = "greater", conf.level= 0.95)  
#making a t test with 'greater than' function to observe if the scores on average will be greater than 100 with a 0.05 significance level
```

```
##  
## One Sample t-test  
##  
## data: problem_2  
## t = -0.59574, df = 24, p-value = 0.7215  
## alternative hypothesis: true mean is greater than 100  
## 95 percent confidence interval:  
## 93.95993 Inf  
## sample estimates:  
## mean of x  
## 98.44
```

*# we fail to reject the null hypothesis due to a p-value of .7215, and we do not suggest that the average IQ scores is not higher than the average IQ 100 score.*

## #Problem 3

### ##Bullet 1

```
expenditure <-  
read.delim("~/GitHub/QT200Spring2020/problem_sets/PS1/expenditure.txt")  
expenditure
```

```
## STATE Y X1 X2 X3 Region  
## 1 ME 61 1704 388 399 1  
## 2 NH 68 1885 372 598 1  
## 3 VT 72 1745 397 370 1  
## 4 MA 72 2394 358 868 1  
## 5 RI 62 1966 357 899 1  
## 6 CT 91 2817 362 690 1  
## 7 NY 104 2685 341 728 1  
## 8 NJ 99 2521 353 826 1  
## 9 PA 70 2127 352 656 1  
## 10 OH 82 2184 387 674 2  
## 11 IN 84 1990 392 568 2  
## 12 IL 84 2435 366 759 2
```

```
## 13 MI 104 2099 403 650 2
## 14 WI 84 1936 393 621 2
## 15 MN 103 1916 402 610 2
## 16 IA 86 1863 385 522 2
## 17 MO 69 2037 364 613 2
## 18 ND 94 1697 429 351 2
## 19 SD 79 1644 411 390 2
## 20 NB 80 1894 379 520 2
## 21 KS 98 2001 380 564 2
## 22 DE 124 2760 388 326 3
## 23 MD 92 2221 393 562 3
## 24 VA 67 1674 402 487 3
## 25 WV 66 1509 405 358 3
## 26 NC 65 1384 423 362 3
## 27 SC 57 1218 453 343 3
## 28 GA 60 1487 420 498 3
## 29 FL 74 1876 334 628 3
## 30 KY 49 1397 594 377 3
## 31 TN 60 1439 346 457 3
## 32 AL 59 1359 637 517 3
## 33 MS 68 1053 448 362 3
## 34 AR 56 1225 403 416 3
## 35 LA 72 1576 433 562 3
## 36 OK 80 1740 378 610 3
## 37 TX 79 1814 409 727 3
## 38 MT 95 1920 412 463 4
## 39 ID 79 1701 418 414 4
## 40 WY 142 2088 415 568 4
## 41 CO 108 2047 399 621 4
## 42 NM 94 1838 458 618 4
## 43 AZ 107 1932 425 699 4
## 44 UT 109 1753 494 665 4
## 45 NV 114 2569 372 663 4
## 46 WA 112 2160 386 584 4
## 47 OR 105 2006 382 534 4
## 48 CA 129 2557 373 717 4
## 49 AK 107 1900 434 379 4
## 50 HI 77 1852 431 693 4
```

## Per capita personal income vs. Per capita expenditure on public education relationship

```
library(tidyverse) #Loading package for graphical usage
```

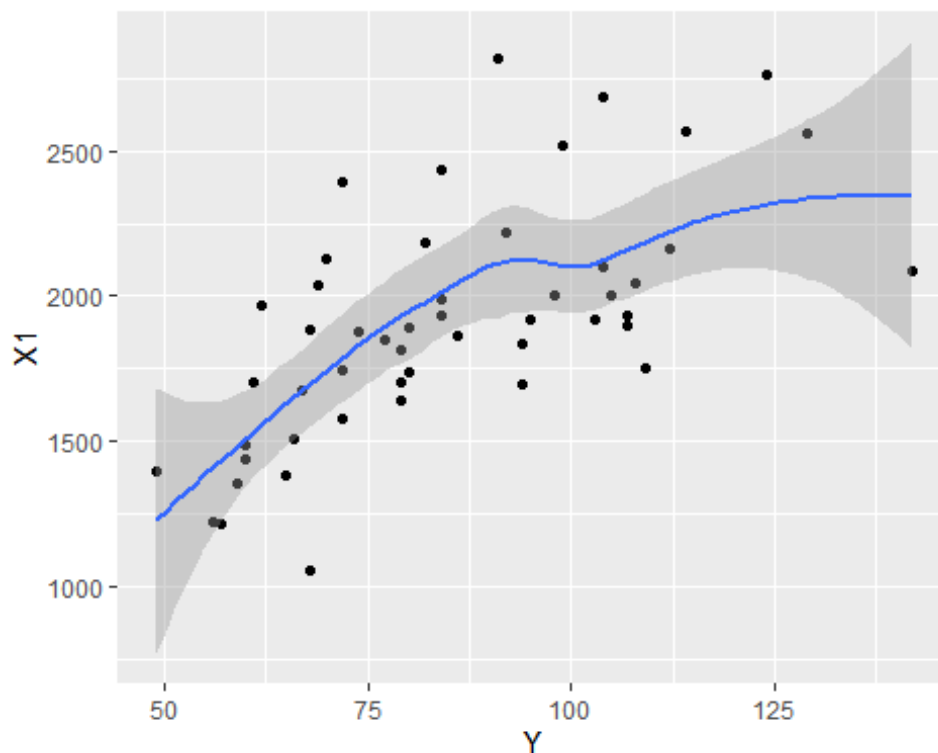
```
## -- Attaching packages -----
----- tidyverse 1.3.0 --

## v ggplot2 3.2.1      v purrr 0.3.3
## v tibble 2.1.3       v dplyr 0.8.3
## v tidyr 1.0.0        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()

ggplot(aes(Y, X1), data=expenditure) + #graphing each axis using variables
within the expenditure data set
  geom_point() + #adding points for each relationship
  geom_smooth() #adding a trendline

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
dev.print(device=pdf, file="Question_3A.pdf")
```

```
## png
## 2
```

```
dev.off()
```

```
## null device
## 1
```

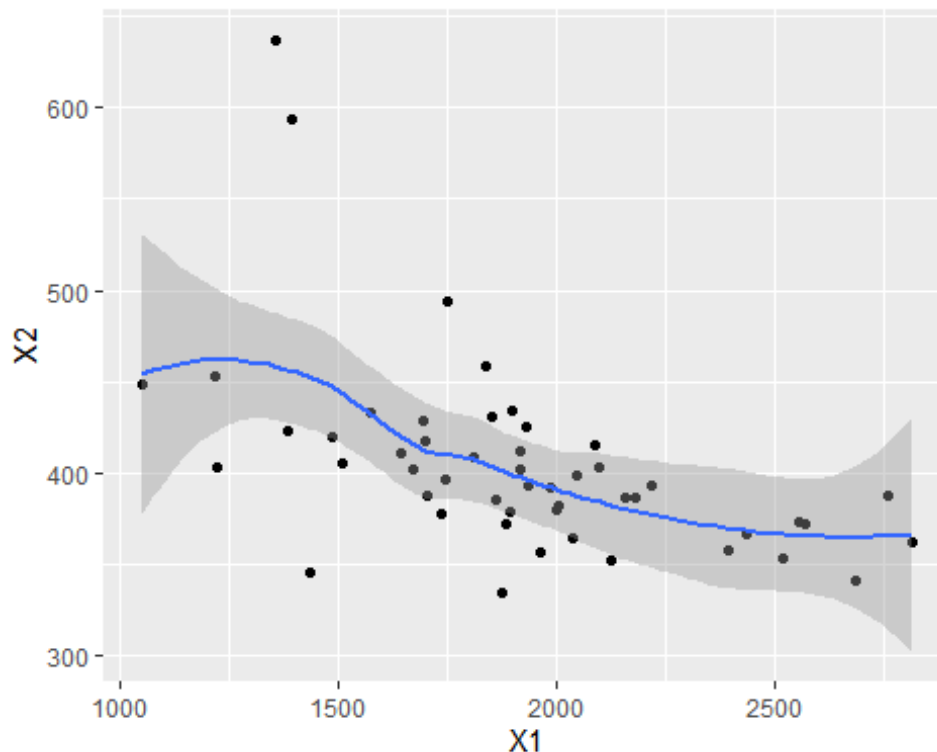
*As per capita expenditure on public education increases, so does the relationship of per capita personal income. This is illustrated by the positive relationship drawn by the trendline. The relationship could be deemed correlational because having more people with more money, leads to more money going back into the system through taxes for example, to be spent on public education. Though one must note, ss per capita expenditure on*

*public education passes approximately 100, the relationship becomes more neutral.*

###Number of residents per thousand under 18 years of age vs. per capita personal income

```
ggplot(aes(X1, X2), data=expenditure) + #graphing each axis using variables
within the expenditure data set
  geom_point() + #adding points for each relationship
  geom_smooth() #adding a trendline

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

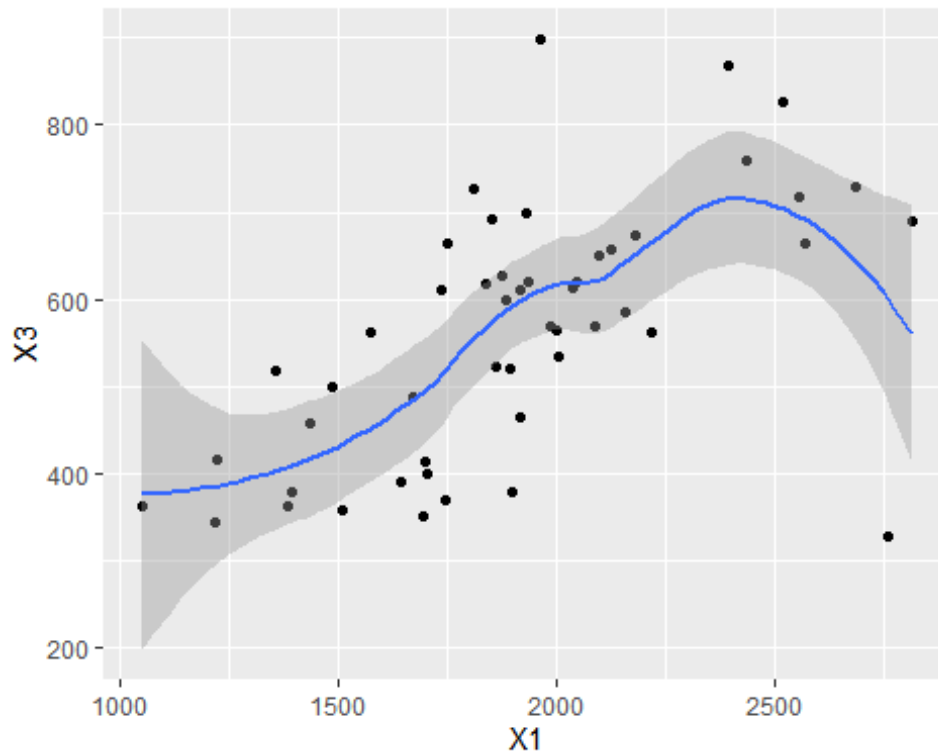


*#As per capita personal income increases, the number of residents per thousand under 18 years of age seem to decrease illustrated by the negative relationship indicated by the trendline. This makes sense, as one is more likely to earn more money later on in their life due to amass amounts of variables such as more educational experience, leading to higher paying jobs.*

**Number of people per thousand residing in urban areas vs. per capita personal income**

```
ggplot(aes(X1, X3), data=expenditure) + #graphing each axis using variables
within the expenditure data set
  geom_point() + #adding points for each relationship
  geom_smooth() #adding a trendline

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



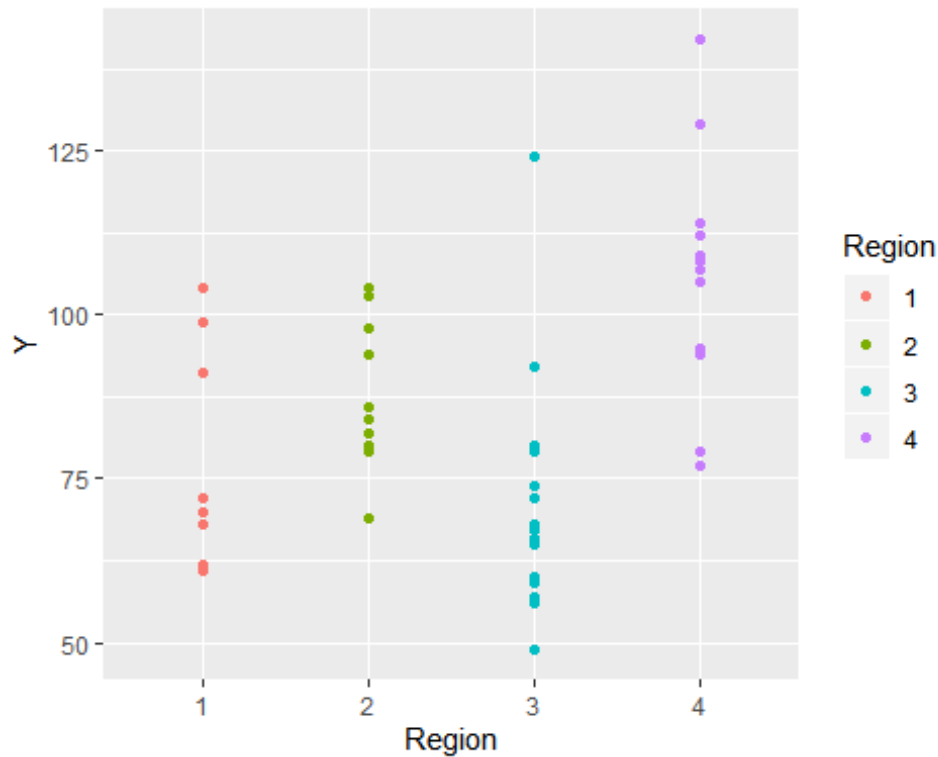
*#As per capita personal income increases, there is a general positive trend of the number of people per thousand residing in urban areas. This could be due to the fact that as personal income increases, one can afford to live in presumably better located residencies in urban areas, closer to work. That relationship because to decrease with the few observations around 2500 on 'X1' possibly because those who gain more money can afford even more luxurious locations outside of the urban areas.*

##Bullet 2

## Per capita expenditure on public education Vs. Region

1 = Northeast | 2 = North Central | 3 = South | 4 = West|

```
expenditure$Region <- as.factor(expenditure$Region)
ggplot(aes(Region, Y), data=expenditure) + #graphing each axis using
variables within the expenditure data set
  geom_point(aes(color = Region)) #adding points for each relationship
```

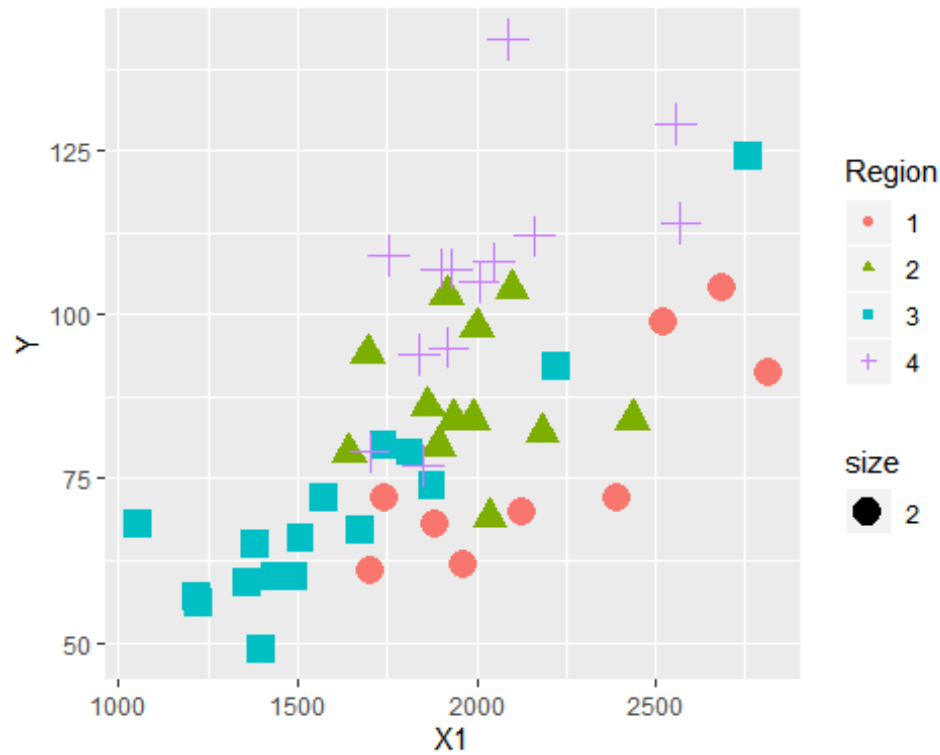


*#For each region, there are differences between the per capita expenditure on public education as indicated by the spread within the graph. On average, the South (Region 3), has the lowest expenditure; whereas the West (Region 4) has the highest expenditure on public education. The North Central region (Region 2) slightly edges out the Northeast (Region 1), but both exist in between the South and West's expenditures*

##Bullet 3

###Per capita personal income vs. Per capita expenditure on public education relationship  
Based on Region

```
ggplot(aes(X1, Y), data=expenditure) + #graphing each axis using variables
within the expenditure data set
  geom_point(aes(color = Region, shape = Region, size=2)) #adding points for
each relationship
```



*#Even when distributing the data by region (as indicated by the different shapes), there is a general positive trend for an increase in per capita expenditure on public education and per capita personal income. This arguments makes sense due to the fact that having more money held by people allows them to give back more money to fund public education through taxes, or even other means.*