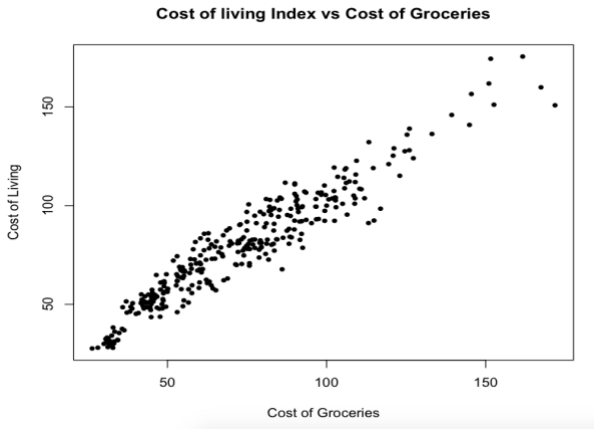
Marcus Crowder

Statistics 3155

R-HW#1

1. Produce a scatterplot between the *Cost of Living Index* and each of the other index variables. As a result, there should be 4 plots in total. Examine the relationship shown in each scatterplot in terms of its form, strength and direction.

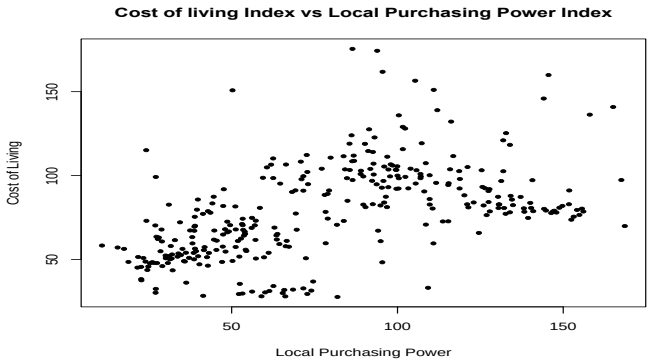


The first plot was comparing Cost of living vs Cost of Groceries. There seems to be a strong positive Association. The form is straight and the direction is rising. It is not a surprise that groceries are more expensive in neighborhoods with a higher cost of living.

The plot between Cost of Living and Cost of Rent is less straightforward. There does seem to be a positive association but the strength seems to be moderate with outliers as points seem to scatter as cost of living increases.

Cost of living vs Restaurant Price Index Appears to have the form of a straight line with a positive Association. It also appears to have a strong association.

Cost of living index vs Purchasing Power Index does not appear to have a form. There seems to be at best a weak positive Association if a line was drawn through it.



All of the R commands used to generate Plots

```
> plot(cost$Rent.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Cost of Groceries", xlab = "Cost of Groceries", ylab = "Cost of Living", pch=20)
> plot(cost$Groceries.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Cost of Groceries", xlab = "Cost of Groceries", ylab = "Cost of Living", pch=20)
> plot(cost$Restaurant.Price.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Restaurant Price", xlab = "Restaurant Price", ylab = "Cost of Living", pch=20)
> plot(cost$Local.Purchasing.Power.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Local Purchasing Power Index", xlab = "Local Purchasing Power", ylab = "Cost of Living",
pch=20)
```

2. Compute the correlations for all the scatterplots obtained above.

```
> cor(cost$Cost.of.Living.Index, cost$Rent.Index)
[1] 0.7722926
> cor(cost$Cost.of.Living.Index, cost$Groceries.Index)
[1] 0.9538616
> cor(cost$Cost.of.Living.Index, cost$Restaurant.Price.Index)
[1] 0.9493554
> cor(cost$Cost.of.Living.Index, cost$Local.Purchasing.Power.Index)
[1] 0.525902
```

3. Verify the conditions for each correlation computed above.

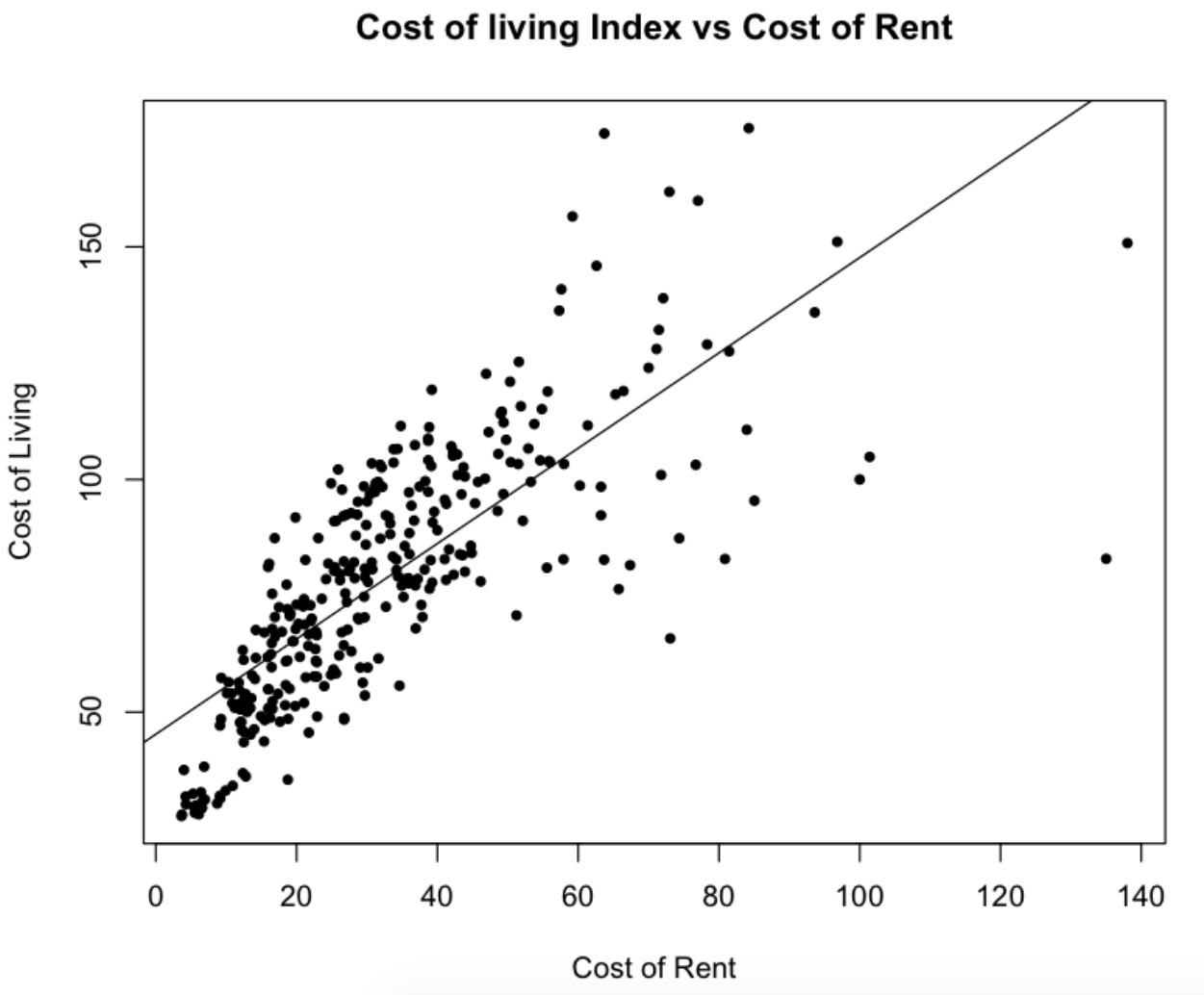
a. The Correlation between Cost of living and Cost of Rent fit the Quantitative condition, it also fits the linearity condition because the straightness of the line is visible. It also satisfies the outlier condition by not having too many outliers.

b. The Correlation between Cost of living and Cost of Groceries fit the Quantitative condition, it also fits the linearity condition the line looks perfectly straight and it has a .95 correlation. There also does not seem to be an outlier

c. The Correlation between Cost of living and Restaurant Prices fit the Quantitative condition, it also fits the linearity condition as the line appears to be straight and there does not seem to be outliers.

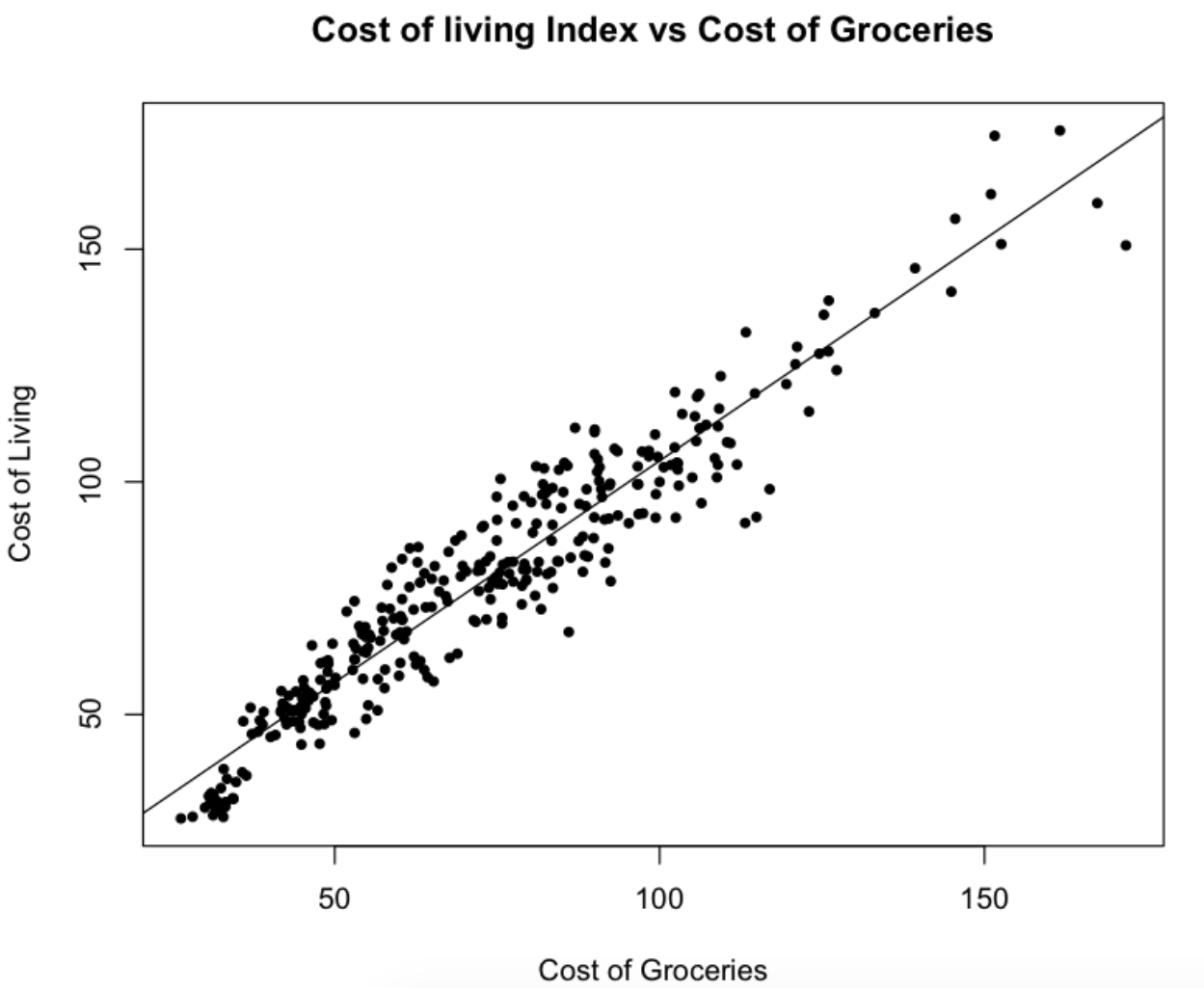
d. The Correlation between Cost of living and Local Purchasing Power fits the Quantitative condition. It barely fits the linearity condition. When I first looked at the scatter plot I assumed they were outliers but just going by the direction of the scatterplot I can't consider those points as outliers and just part of the cost of living behavior.

Fit a regression line through models



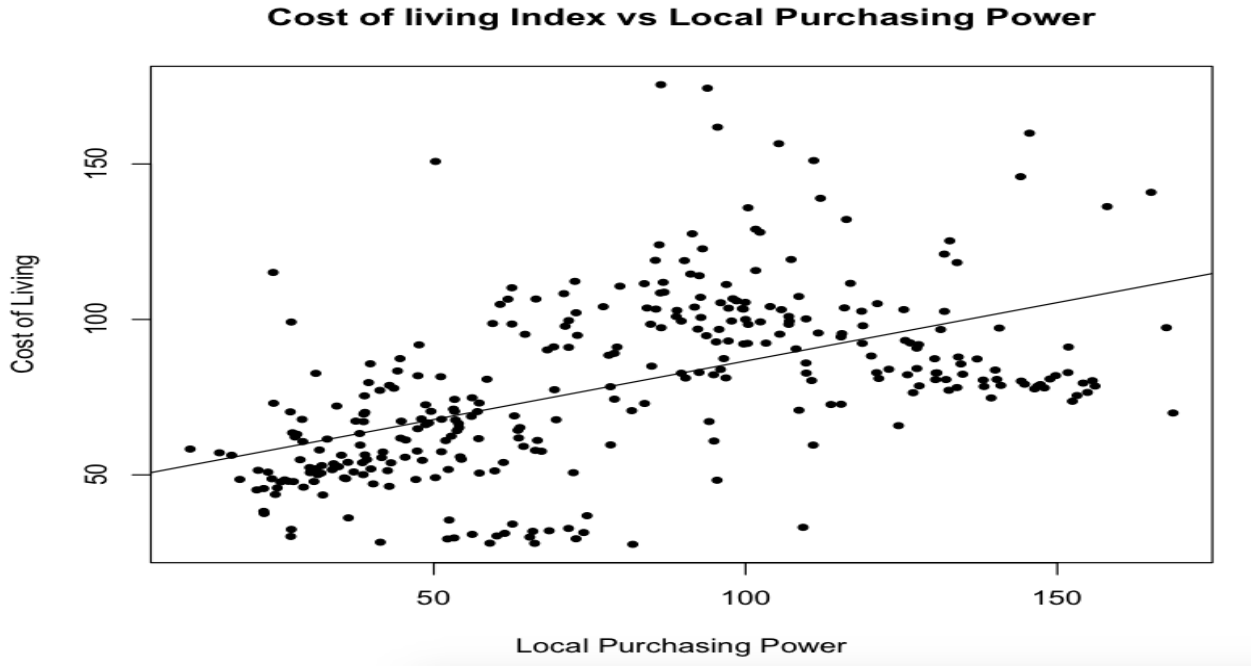

```
> imod <- lm(Cost.of.Living.Index ~ Rent.Index, data = cost)
> summary(imod)
> plot(cost$Rent.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Cost
of Rent", xlab = "Cost of Rent", ylab = "Cost of Living", pch=20)
> abline(imod)
```

high intercept slope close to 1




```
> plot(cost$Rent.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Cost
of Rent", xlab = "Cost of Rent", ylab = "Cost of Living", pch=20)
> abline(imod)
> imod <- lm(Cost.of.Living.Index ~ Groceries.Index, data = cost)
> plot(cost$Groceries.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs
Cost of Groceries", xlab = "Cost of Groceries", ylab = "Cost of Living", pch=20)
> abline(imod)
```


```
imod <- lm(Cost.of.Living.Index ~ Restaurant.Price.Index, data = cost)
plot(cost$Restaurant.Price.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs
Restaurant Price", xlab = "Restaurant Price", ylab = "Cost of Living", pch=20)
> abline(imod)
```



Cost of living has a small slope

```
imod <- lm(Cost.of.Living.Index ~ Local.Purchasing.Power.Index, data = cost)
> plot(cost$Local.Purchasing.Power.Index, cost$Cost.of.Living.Index, main = "Cost of living Index vs Local Purchasing Power", xlab = "Local Purchasing Power", ylab = "Cost of Living", pch=20)
> abline(imod)
```

R commands for Linear Models.

```
> imod <- lm(Cost.of.Living.Index ~ Groceries.Index, data = cost)
> summary(imod)

Call:
lm(formula = Cost.of.Living.Index ~ Groceries.Index, data = cost)

Residuals:
    Min      1Q  Median      3Q     Max
-26.2714  -6.2766   0.4478   5.2780  20.7336

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.21784    1.31039   7.034 1.22e-11 ***
Groceries.Index  0.95295    0.01677  56.831  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.356 on 320 degrees of freedom
Multiple R-squared:  0.9099,  Adjusted R-squared:  0.9096
F-statistic:  3230 on 1 and 320 DF,  p-value: < 2.2e-16

> imod <- lm(Cost.of.Living.Index ~ Rent.Index, data = cost)
> summary(imod)

Call:
lm(formula = Cost.of.Living.Index ~ Rent.Index, data = cost)

Residuals:
    Min      1Q  Median      3Q     Max
-100.617  -10.622    0.012   10.353   63.869

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 45.23260    1.83225   24.69   <2e-16 ***
Rent.Index   1.02462    0.04712   21.75   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.68 on 320 degrees of freedom
Multiple R-squared:  0.5964,  Adjusted R-squared:  0.5952
F-statistic: 472.9 on 1 and 320 DF,  p-value: < 2.2e-16

> imod <- lm(Cost.of.Living.Index ~ Restaurant.Price.Index, data = cost)
> summary(imod)

Call:
lm(formula = Cost.of.Living.Index ~ Restaurant.Price.Index, data = cost)

Residuals:
   Min     1Q  Median     3Q    Max
-26.638  -5.380  -0.786   5.153  37.062

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          24.66360    1.11432   22.13   <2e-16 ***
Restaurant.Price.Index  0.80333    0.01486   54.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.744 on 320 degrees of freedom
Multiple R-squared:  0.9013,  Adjusted R-squared:  0.901
F-statistic:  2921 on 1 and 320 DF,  p-value: < 2.2e-16

> imod <- lm(Cost.of.Living.Index ~ Local.Purchasing.Power.Index, data = cost)
> summary(imod)

Call:
lm(formula = Cost.of.Living.Index ~ Local.Purchasing.Power.Index,
   data = cost)

Residuals:
   Min     1Q  Median     3Q    Max
-56.956 -14.145  -2.834  14.212  94.005

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)                  48.99742    3.00227   16.32   <2e-16 ***
Local.Purchasing.Power.Index  0.37616    0.03401   11.06   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.67 on 320 degrees of freedom
Multiple R-squared:  0.2766,  Adjusted R-squared:  0.2743
F-statistic: 122.3 on 1 and 320 DF,  p-value: < 2.2e-16
```

5.The best predictor for cost of living would have to be The Price of Groceries. It is shown to have the highest correlation. It also has a lower standard error and residual error than the Cost of Rent the 2nd leading candidate. My assumptions were based on the scatter plots and summary statistics above. The worst indicator would local purchasing power. It has the lowest correlation and the points on the scatter plot look like it has a weak association.

6.

```
> imod <- lm(Cost.of.Living.Index ~ Groceries.Index, data = cost)
> cost[172,]
            City Cost.of.Living.Index Rent.Index Groceries.Index Restaurant.Price.Index
Local.Purchasing.Power.Index
172 Beijing, China               77.19      34.93          83.57                 43.61
41.35
> imod$fitted.values[172]
    172
88.85556
> imod$residuals[172]
     172
-11.66556
```

The fitted value is 88.85556

And the residuals is
11.66556