I. INTRODUCTION

Nowadays, all over the world have their own data collection to store and analyze important information. For this project, I was interested to study about age-adjusted death rates for selected causes of death for all ages, by sex in United States (US) from 2007 until 2017 because US is the third largest population in the world with 318,816,000 in 2020. In US, there have 6 popular disease causes of death which are Heart disease, Cancer, Unintentional injuries, CLRD (Chronic Lower Respiratory disease), Stroke, Diabetes, Alzheimer's disease. According the data, I want to find out more either there has the relationship between all the disease and gender.

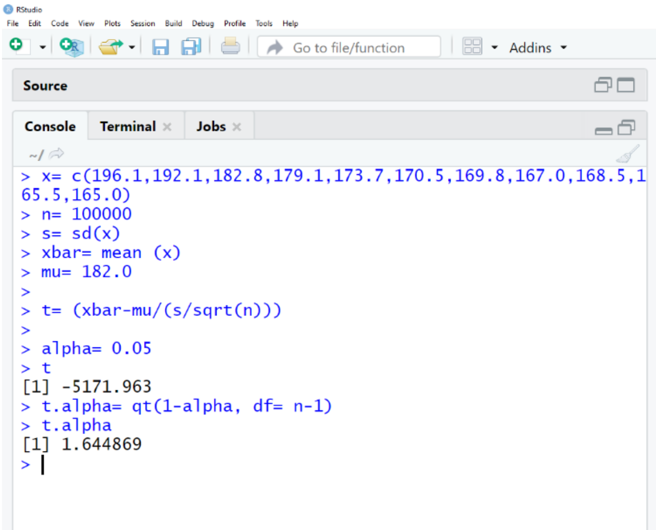
II. METHODOLOGY

This is the parameter and variables from the data collected:

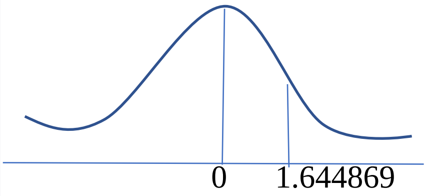
Data collected	**Level of measurement**	**Description**
Gender	Interval	Male, Female
Year	Interval	2007-2017
Heart disease	Ratio	-
Cancer	Ratio	-
Unintentional injuries	Ratio	-
CLRD (Chronic Lower Respiratory disease)	Ratio	-
Stroke	Ratio	-
Diabetes	Ratio	-
Alzheimer's disease	Ratio	-

III. RESULT AND DISCUSSION

1 sample test (Population mean with unknown variance- one tailed)



```
> x= c(196.1,192.1,182.8,179.1,173.7,170.5,169.8,167.0,168.5,1
65.5,165.0)
> n= 100000
> s= sd(x)
> xbar= mean (x)
> mu= 182.0
>
> t= (xbar-mu/(s/sqrt(n)))
>
> alpha= 0.05
> t
[1] -5171.963
> t.alpha= qt(1-alpha, df= n-1)
> t.alpha
[1] 1.644869
> |
```



H_0: μ=182

H_1: μ > 182

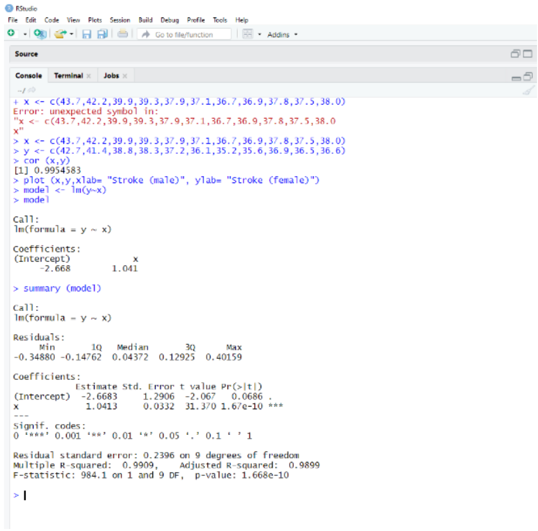
Test statistic= -5171.963

critical value = 1.644869

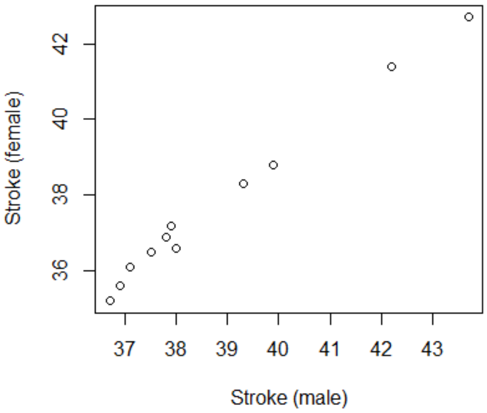
Decision: Fail to reject H_0 since critical value greater than t statistic, 1.644869>-5171. 963.

Conclusion: There is mean no coefficient of restitution exceeds 182.0 at 0.05 significance level.

Correlation between female had stroke and male had stroke



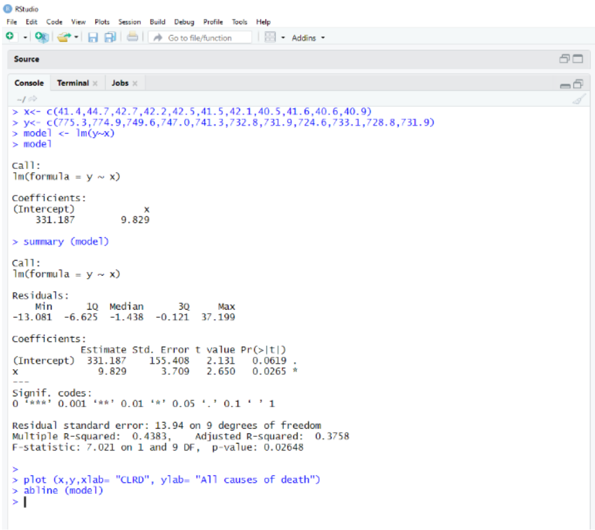
```
+ x <- c(43.7,42.2,39.9,39.3,37.9,37.1,36.7,36.9,37.8,37.5,38.0)
Error: unexpected symbol in:
"x <- c(43.7,42.2,39.9,39.3,37.9,37.1,36.7,36.9,37.8,37.5,38.0
x"
> x <- c(43.7,42.2,39.9,39.3,37.9,37.1,36.7,36.9,37.8,37.5,38.0)
> y <- c(42.7,41.4,38.8,38.3,37.2,36.1,35.2,35.6,36.9,36.5,36.6)
> cor (x,y)
[1] 0.9954583
> plot (x,y,xlab= "Stroke (male)", ylab= "Stroke (female)")
> model <- lm(y~x)
> model

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
     -2.668        1.041

> summary (model)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-0.34880 -0.14762  0.04372  0.12925  0.40159

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.6683     1.2906  -2.067   0.0686 .
x             1.0413     0.0332  31.370 1.67e-10 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2396 on 9 degrees of freedom
Multiple R-squared:  0.9909,	Adjusted R-squared:  0.9899
F-statistic: 984.1 on 1 and 9 DF,  p-value: 1.668e-10

> |
```



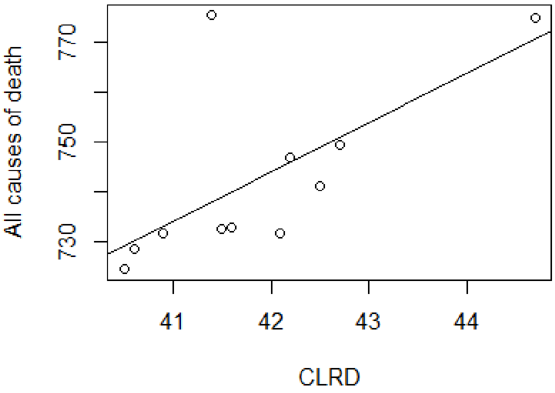
Stroke (male) (x)	43.7	42.2	39.9	39.3	37.9	37.1	36.7	36.9	37.8	37.5	38.0
Stroke (female) (y)	42.7	41.4	38.8	38.3	37.2	36.1	35.2	35.6	36.9	36.5	36.6

Based from scatter plot, there is linear relationship between female had stroke and male had stroke. The relationship is positive strong correlation because of value r is 0.9954583 . The value r is very consistance. When the number of male had stroke increase, female had stroke also increase.

Regression between CLRD (Chronic Lower Respiratory disease) and all causes of death



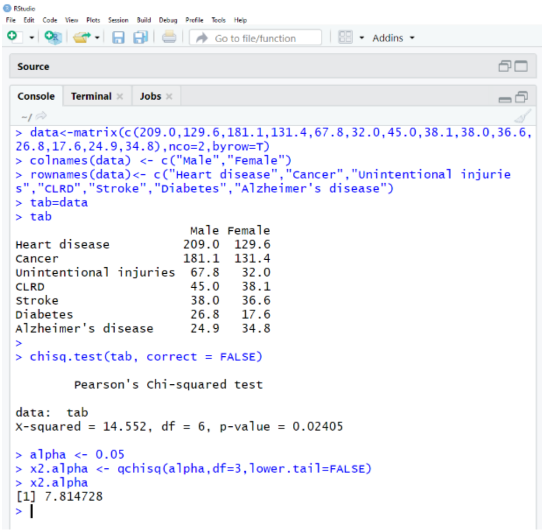
```
> x<- c(41.4,44.7,42.7,42.2,42.5,41.5,42.1,40.5,41.6,40.6,40.9)
> y<- c(775.3,774.9,749.6,747.0,741.3,732.8,731.9,724.6,733.1,728.8,731.9)
> model <- lm(y~x)
> model

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    331.187        9.829

> summary (model)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-13.081  -6.625  -1.438  -0.121  37.199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  331.187    155.408   2.131   0.0619 .
x              9.829      3.709   2.650   0.0265 *
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.94 on 9 degrees of freedom
Multiple R-squared:  0.4383,    Adjusted R-squared:  0.3758
F-statistic: 7.021 on 1 and 9 DF,  p-value: 0.02648

>
> plot (x,y,xlab= "CLRD", ylab= "All causes of death")
> abline (model)
> |
```



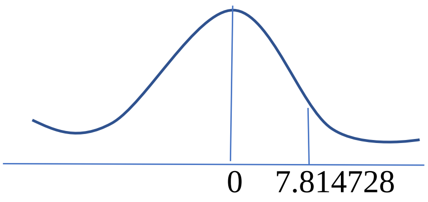
CLRD (x)	36.6	39.8	38.3	38.0	38.5	37.8	38.5	37.1	38.6	37.4	38.1
All causes of death (y)	658.1	659.1	636.8	634.9	632.4	624.7	623.5	616.7	624.2	617.5	619.7

Based from scatter plot, there is weaker linear relationship between CLRD (Chronic Lower Respiratory disease) and all causes of death because r^2= 0.4383. The regression equation is ŷ= 331.187 + 9.829x.

Pearson’s Chi square test of independence between gender and causes of death (two ways)



```
> data<-matrix(c(209.0,129.6,181.1,131.4,67.8,32.0,45.0,38.1,38.0,36.6,
26.8,17.6,24.9,34.8),nco=2,byrow=T)
> colnames(data) <- c("Male","Female")
> rownames(data)<- c("Heart disease","Cancer","Unintentional injurie
s","CLRD","Stroke","Diabetes","Alzheimer's disease")
> tab=data
> tab
                        Male Female
Heart disease          209.0  129.6
Cancer                 181.1  131.4
Unintentional injuries  67.8   32.0
CLRD                    45.0   38.1
Stroke                  38.0   36.6
Diabetes                26.8   17.6
Alzheimer's disease     24.9   34.8
>
> chisq.test(tab, correct = FALSE)

        Pearson's Chi-squared test

data:  tab
X-squared = 14.552, df = 6, p-value = 0.02405

> alpha <- 0.05
> x2.alpha <- qchisq(alpha,df=3,lower.tail=FALSE)
> x2.alpha
[1] 7.814728
> |
```



H₀: gender is independent of causes of death.

H₁: gender is not independent of causes of death.

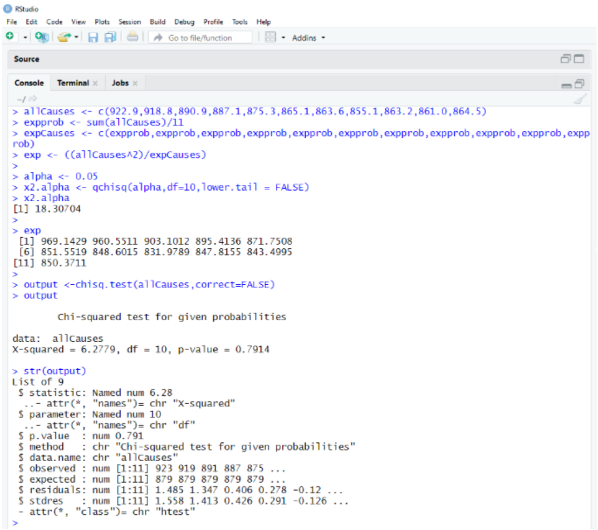
Test statistic = 14.522

Critical value = 7.814728.

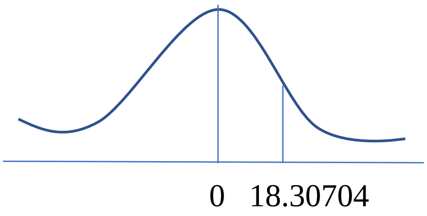
Decision: Since critical value is less than test statistic (7.814728<14.522), thus reject H₀.

Conclusion: There is not enough evidence that gender is independent of causes of death at 0.05 significance level.

Chi square test of independence between all causes (one way)



```
> allCauses <- c(922.9,918.8,890.9,887.1,875.3,865.1,863.6,855.1,863.2,861.0,864.5)
> expprob <- sum(allCauses)/11
> expCauses <- c(expprob,expprob,expprob,expprob,expprob,expprob,expprob,expprob,expprob,expprob,expp
rob)
> exp <- ((allCauses^2)/expCauses)
>
> alpha <- 0.05
> x2.alpha <- qchisq(alpha,df=10,lower.tail = FALSE)
> x2.alpha
[1] 18.30704
>
> exp
 [1] 969.1429 960.5511 903.1012 895.4136 871.7508
 [6] 851.5519 848.6015 831.9789 847.8155 843.4995
[11] 850.3711
>
> output <-chisq.test(allCauses,correct=FALSE)
> output

	Chi-squared test for given probabilities

data:  allCauses
X-squared = 6.2779, df = 10, p-value = 0.7914

> str(output)
List of 9
 $ statistic: Named num 6.28
  ..- attr(*, "names")= chr "X-squared"
 $ parameter: Named num 10
  ..- attr(*, "names")= chr "df"
 $ p.value  : num 0.791
 $ method   : chr "Chi-squared test for given probabilities"
 $ data.name: chr "allCauses"
 $ observed : num [1:11] 923 919 891 887 875 ...
 $ expected : num [1:11] 879 879 879 879 879 ...
 $ residuals: num [1:11] 1.485 1.347 0.406 0.278 -0.12 ...
 $ stdres   : num [1:11] 1.558 1.413 0.426 0.291 -0.126 ...
 - attr(*, "class")= chr "htest"
>
```



H₀: $p_1=p_2=p_3=p_4=p_5=p_6=p_7=p_8$

H₁: at least one is different.

Test statistic = 6.2779

Critical value = 18.30704

Decision: Since critical value is greater than test statistic (18.30704<6.2779), thus fail to reject H₀.

Conclusion: There is not enough evidence of a differences in all causes of death from 2007-2017 at 0.05 significance level.

IV. Conclusion

Based on data, the sample population was taken for 100000 people in US. For 1 sample test, I was tested on population mean =182. I can conclude that fail to reject because there is mean no coefficient of restitution exceeds 182.0 at 0.05 significance level. Next**,** correlation between female had stroke and male had stroke have positive strong correlation. When the number of male had stroke increase, female had stroke also increase. While for regression between CLRD (Chronic Lower Respiratory disease) and all causes of death have linear relationship. Pearson's Chi square test of independence between gender and causes of death (two ways), I was tested on gender is independent of causes of death. From the value of critical value is less than test statistic, I can conclude that I need to reject H_0 because there is not enough evidence that gender is independent of causes of death at 0.05 significance level. Lastly, Chi square test of independence between all causes (one way)**.** Since critical value is greater than test statistic (18.30704<6.2779), thus fail to reject H_0. There is not enough evidence of a differences in all causes of death from 2007-2017 at 0.05 significance level.