Econ4274_problem_set_2

Man Adrian Ho Hin

2023-03-14

```
rm(list=ls())
setwd("/Users/adrian/Desktop/econ4274_problem_set_2")
```

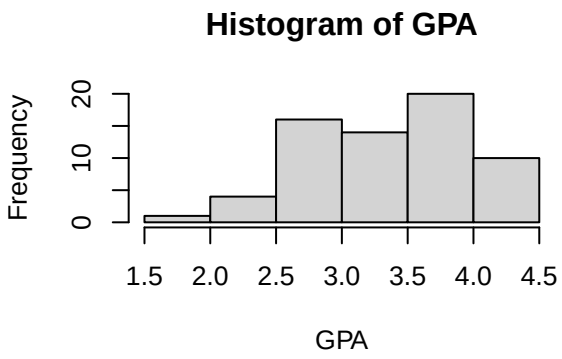
Question 1a

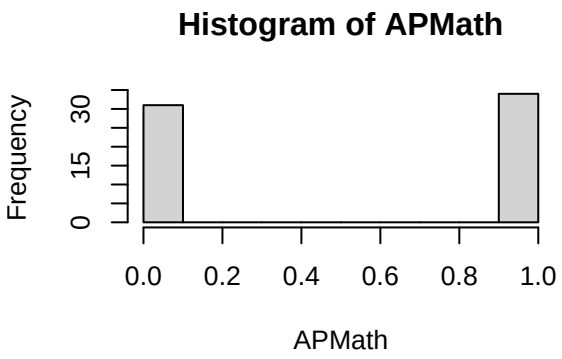
```
sat = read.csv("SAT.csv")
summary(sat)
```

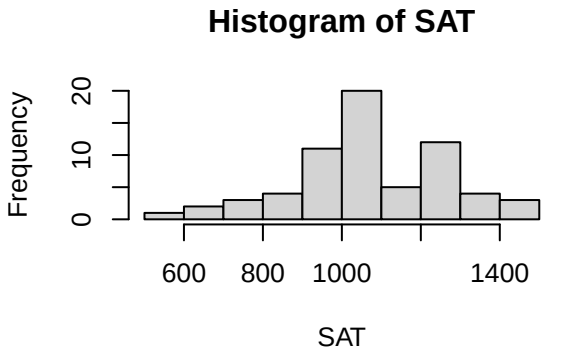
```
##       obs             AP              APEng            APMath            ESL
##  Min.   : 1   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0
##  1st Qu.:17   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0
##  Median :33   Median :1.0000   Median :1.0000   Median :1.0000   Median :0.0
##  Mean   :33   Mean   :0.6769   Mean   :0.5538   Mean   :0.5231   Mean   :0.4
##  3rd Qu.:49   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0
##  Max.   :65   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0
##      gender            GPA             prep             race
##  Min.   :0.0000   Min.   :1.640   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:2.940   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :3.470   Median :1.0000   Median :0.0000
##  Mean   :0.4923   Mean   :3.362   Mean   :0.7385   Mean   :0.3231
##  3rd Qu.:1.0000   3rd Qu.:3.870   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :4.380   Max.   :1.0000   Max.   :1.0000
##       SAT
##  Min.   : 590
##  1st Qu.: 970
##  Median :1070
##  Mean   :1076
##  3rd Qu.:1230
##  Max.   :1430
```

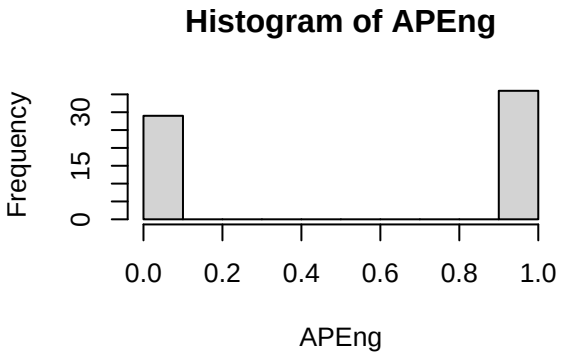
Question 1b

```
par(mfcol = c(2, 2))
with(sat, hist(GPA))
with(sat, hist(SAT))
with(sat, hist(APMath))
with(sat, hist(APEng))
```



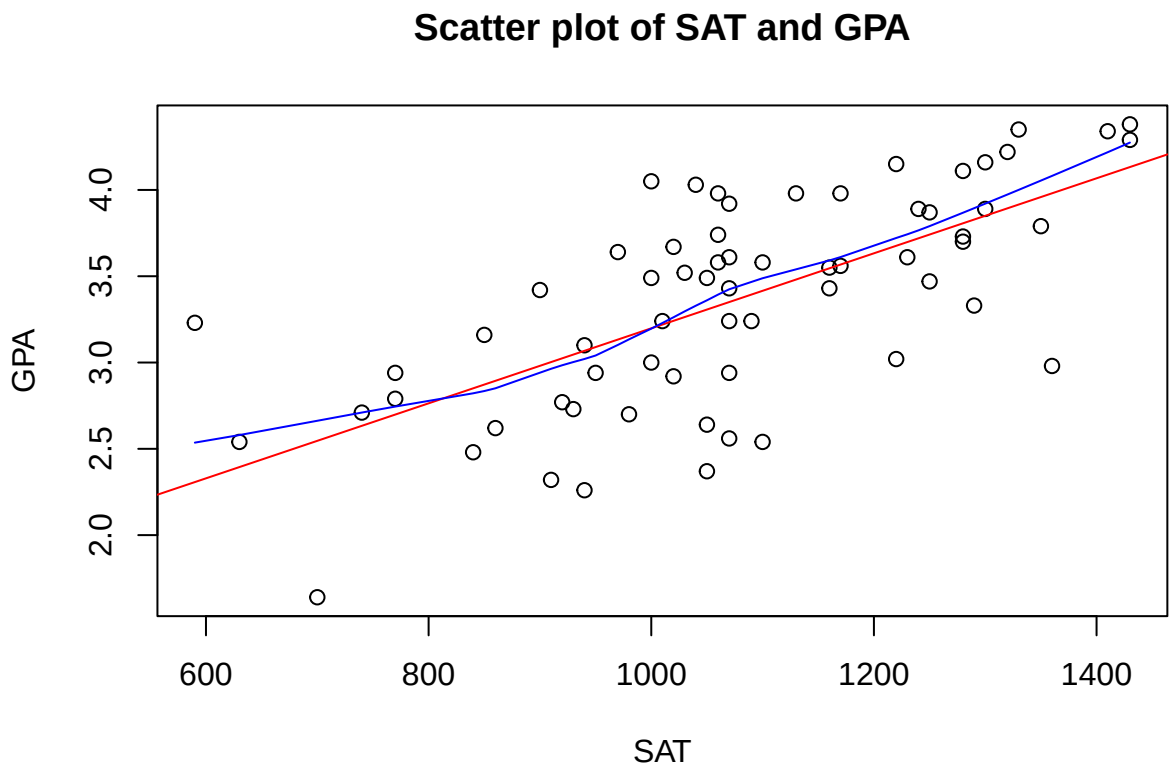






Question 1c

```
with(sat, plot(SAT, GPA, main = "Scatter plot of SAT and GPA"))
abline(lm(GPA~SAT, data = sat), col = "red")
lines(lowess(sat$SAT, sat$GPA), col="blue")
```



Question 1d

```
# i
x_1 = rep(1,length(sat$SAT))

# ii
X = with(sat, matrix(c(x_1,SAT,APMath,APEng,ESL,gender,race),nrow=length(SAT)))

# iii
y = sat$GPA

# iv
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y
beta_hat = t(beta_hat)
colnames(beta_hat)=c("Intercept","SAT","APMath","APEng","ESL","gender","race")

# v
lm1 = with(sat, lm(GPA~SAT+APMath+APEng+ESL+gender+race))
cat("By matrix algebra: \n")
```

```
## By matrix algebra:
```

```
print(beta_hat)
```

```
##      Intercept         SAT    APMath     APEng         ESL     gender
## [1,]   1.66719 0.001319036 0.1408304 0.5032422 -0.07609953 -0.1477924
##            race
## [1,] 0.08437138
```

```
cat("\n")
```

```
cat("By lm() function: \n")
```

```
## By lm() function:
```

```
print(lm1$coefficients)
```

```
##  (Intercept)          SAT       APMath        APEng          ESL       gender
##  1.667190138  0.001319036  0.140830421  0.503242246 -0.076099529 -0.147792393
##         race
##  0.084371380
```

Question 2

```
rm(list=ls())
drug_price = read.csv("drug_price.csv")

lm1 = with(drug_price, lm(p.r~GDP.r))
lm2 = with(drug_price, lm(p.r~GDP.r+cv+cv.r))
lm3 = with(drug_price, lm(p.r~GDP.r+p.control+p.comp))
lm4 = with(drug_price, lm(p.r~GDP.r+patent))
lm5 = with(drug_price, lm(p.r~GDP.r+cv+cv.r+p.control+p.comp+patent))

library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(lm1,lm2,lm3,lm4,lm5,type="text")
```

```
##
## ===============================================================================================
##                                                        Dependent variable:
##                     ------------------------------------------------------------------------
##                                                              p.r
```

```
##                                (1)                    (2)                    (3)                    (4)
## ------------------------------------------------------------------------------------------------------
## GDP.r                       0.923***               1.650***               0.974***               0.87
##                             (0.144)                (0.249)                (0.118)                (0.1
##
## cv                                                  -0.111
##                                                    (0.247)
##
## cv.r                                               -0.822***
##                                                    (0.283)
##
## p.control                                                                -21.518***
##                                                                           (7.402)
##
## p.comp                                                                    -14.386*
##                                                                           (7.865)
##
## patent                                                                                           13.9
##                                                                                                  (8.0
##
## Constant                    34.497***              32.231***              44.427***              29.72
##                             (7.390)                (6.453)                (6.401)                (7.6
##
## ------------------------------------------------------------------------------------------------------
## Observations                  32                     32                     32                     3
## R2                          0.577                  0.706                  0.747                  0.6
## Adjusted R2                 0.563                  0.675                  0.720                  0.5
## Residual Std. Error    22.972 (df = 30)       19.819 (df = 28)       18.382 (df = 28)       22.239 (
## F Statistic         40.936*** (df = 1; 30) 22.435*** (df = 3; 28) 27.596*** (df = 3; 28) 23.344*** (
## ======================================================================================================
## Note:
```

```
anova(lm3,lm5)
```

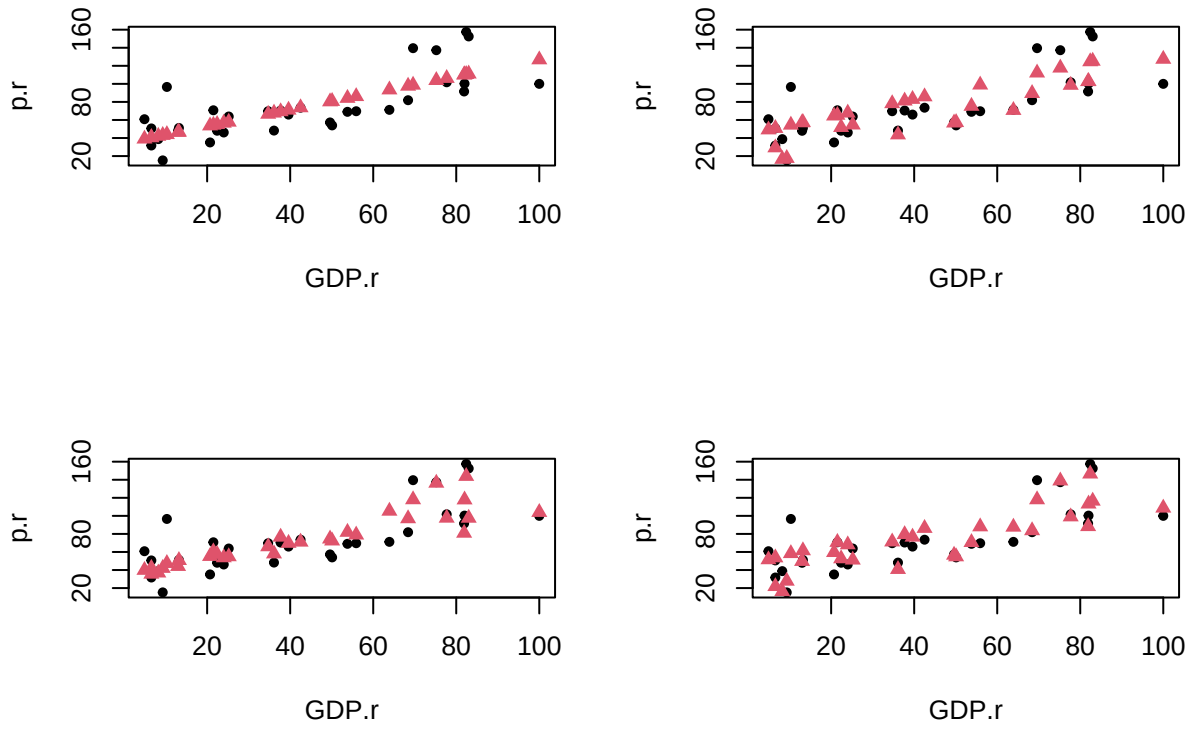
```
## Analysis of Variance Table
##
## Model 1: p.r ~ GDP.r + p.control + p.comp
## Model 2: p.r ~ GDP.r + cv + cv.r + p.control + p.comp + patent
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     28 9461.2
## 2     25 6864.8  3    2596.4 3.1518 0.04256 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfcol = c(2, 2))
plot(p.r~GDP.r,data=drug_price,pch=20)
points(drug_price$GDP.r,lm1$fitted.values,pch=17,col=2)

plot(p.r~GDP.r,data=drug_price,pch=20)
points(drug_price$GDP.r,lm2$fitted.values,pch=17,col=2)

plot(p.r~GDP.r,data=drug_price,pch=20)
points(drug_price$GDP.r,lm3$fitted.values,pch=17,col=2)
```

```
plot(p.r~GDP.r,data=drug_price,pch=20)
points(drug_price$GDP.r,lm5$fitted.values,pch=17,col=2)
```



Question 3a

```
rm(list=ls())
tsssu = read.csv("TSSSU.csv")
summary(tsssu)
```

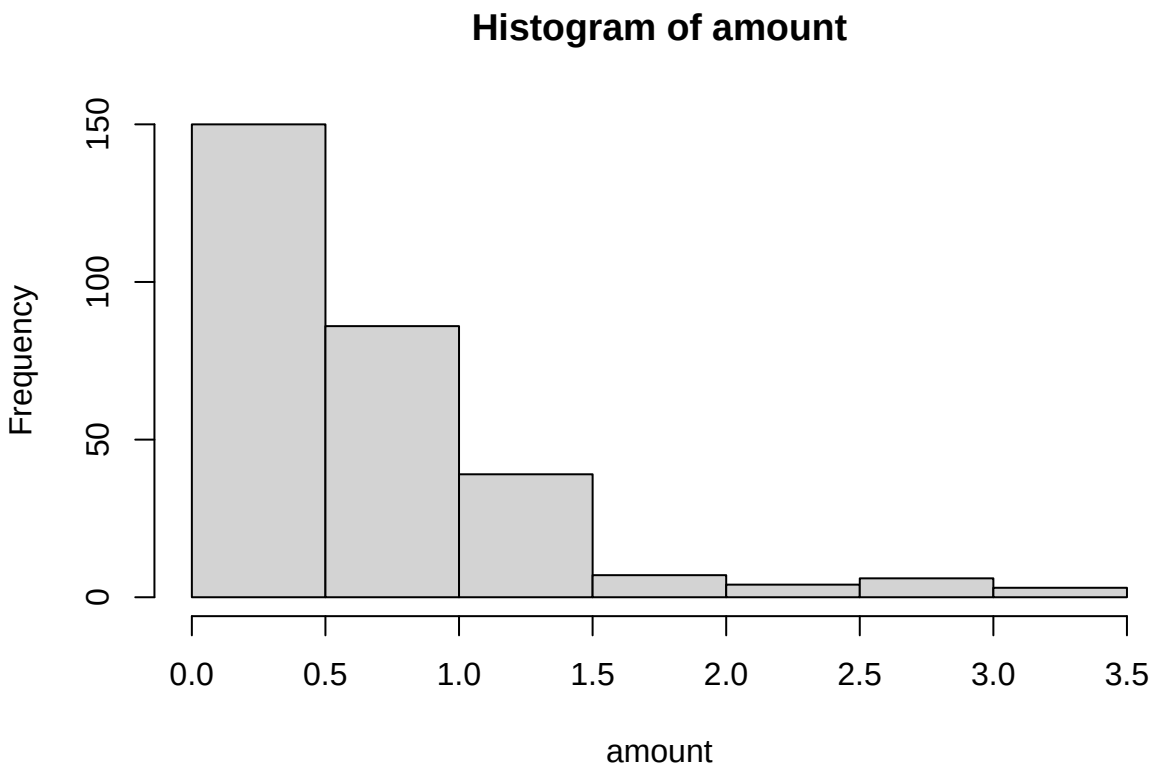
```
##        id            name              amount         abstract
##  Min.   :  1.0   Length:295         Min.   :0.1000   Length:295
##  1st Qu.: 74.5   Class :character   1st Qu.:0.3000   Class :character
##  Median :148.0   Mode  :character   Median :0.5000   Mode  :character
##  Mean   :148.0                      Mean   :0.7078
##  3rd Qu.:221.5                      3rd Qu.:1.0000
##  Max.   :295.0                      Max.   :3.2800
##
##    address           ref_number         university            year
##  Length:295         Length:295         Length:295         Min.   :2014
##  Class :character   Class :character   Class :character   1st Qu.:2015
##  Mode  :character   Mode  :character   Mode  :character   Median :2017
```

```
##                                                        Mean   :2017
##                                                        3rd Qu.:2019
##                                                        Max.   :2020
##
##      area              survive        social_media      phone_call
##  Length:295         Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :1.0000   Median :0.0000   Median :0.0000
##                     Mean   :0.8441   Mean   :0.3299   Mean   :0.1565
##                     3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##                                      NA's   :1        NA's   :1
##    No_members       No_alumni     No_professor     No_undergrad
##  Min.   : 0.000   Min.   :0.0   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 2.000   1st Qu.:0.5   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 3.000   Median :1.0   Median :1.0000   Median :0.0000
##  Mean   : 3.763   Mean   :1.4   Mean   :0.7085   Mean   :0.2169
##  3rd Qu.: 5.000   3rd Qu.:2.0   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :15.000   Max.   :8.0   Max.   :5.0000   Max.   :4.0000
##
##   No_postgrad         No_other          Employee       private_fund
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 1.000   Min.   :0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 2.000   1st Qu.:0.0000
##  Median : 0.0000   Median : 0.0000   Median : 4.000   Median :0.0000
##  Mean   : 0.5186   Mean   : 0.9186   Mean   : 6.034   Mean   :0.1254
##  3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 7.000   3rd Qu.:0.0000
##  Max.   :10.0000   Max.   :11.0000   Max.   :51.000   Max.   :1.0000
##                                      NA's   :207
##   Sciencepark
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3627
##  3rd Qu.:1.0000
##  Max.   :1.0000
##
```

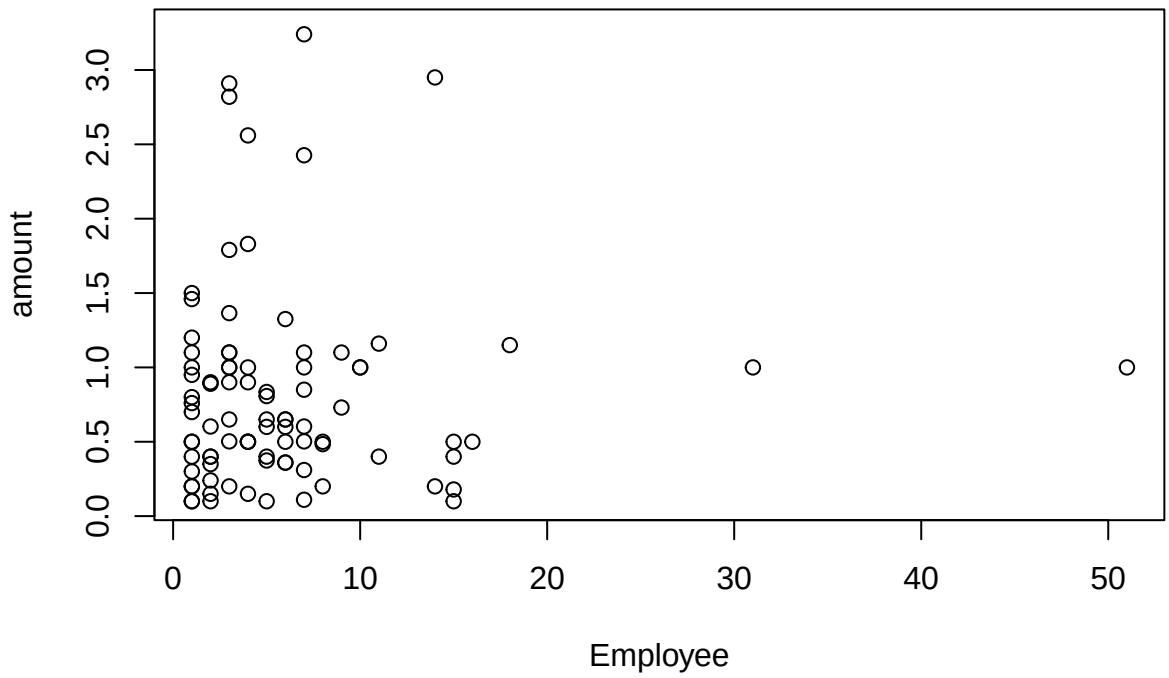
```
#num_cols <- unlist(lapply(tsssu, is.numeric))
#summary(tsssu[ , num_cols])
```

Question 3b

```
#tsssu$Employee[is.na(tsssu$Employee)]=6
with(tsssu, hist(amount))
```




```
with(tsssu, plot(Employee, amount))
```



Question 3c

```
library(stargazer)

tsssu$f.university=as.factor(tsssu$university)
tsssu$f.year=as.factor(tsssu$year)
tsssu$f.area=as.factor(tsssu$area)

m1 = lm(survive~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsssu)
m2 = lm(Employee~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsssu)
m3 = lm(social_media~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsss
m4 = lm(phone_call~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsssu
m5 = lm(private_fund~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsss
m6 = lm(Sciencepark~amount+No_undergrad+No_postgrad+No_professor+f.university+f.year+f.area,data = tsssu

# Further modify in the excel
stargazer(m1,m2,m3,m4,m5,m6,type="html",out="regtab.html")
```

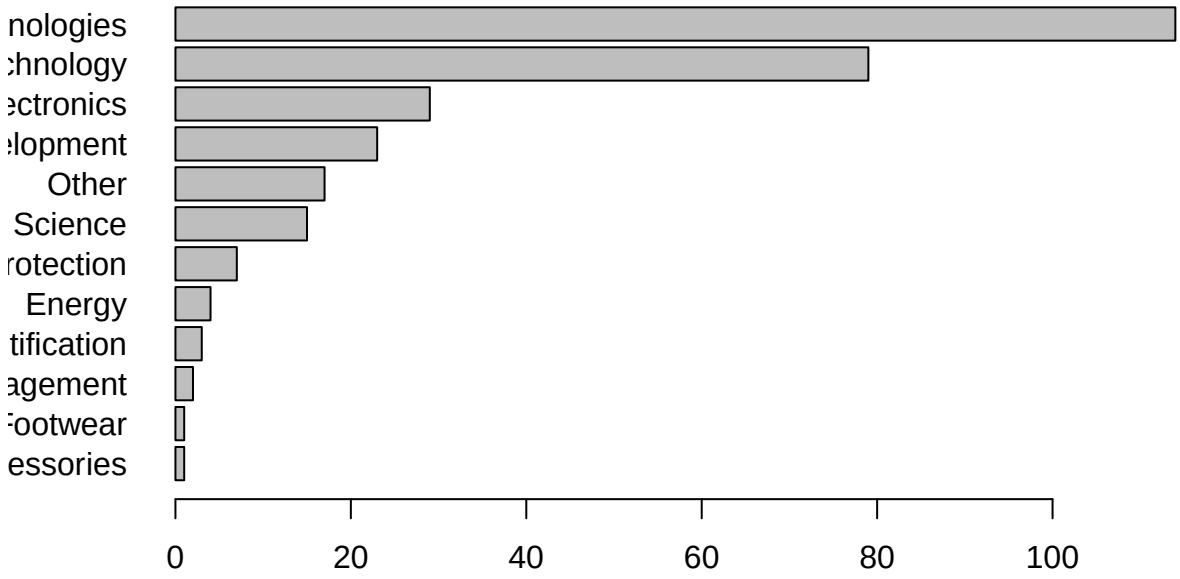
```
##
## <table style="text-align:center"><tr><td colspan="7" style="border-bottom: 1px solid black"></td></t
## <tr><td></td><td colspan="6" style="border-bottom: 1px solid black"></td></tr>
## <tr><td style="text-align:left"></td><td>survive</td><td>Employee</td><td>social_media</td><td>phone_
## <tr><td style="text-align:left"></td><td>(1)</td><td>(2)</td><td>(3)</td><td>(4)</td><td>(5)</td><td
```

```
## <tr><td colspan="7" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left"></td><td>(0.049)</td><td>(1.887)</td><td>(0.064)</td><td>(0.051)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">No_undergrad</td><td>0.012</td><td>-0.801</td><td>-0.013</td><td>0.03
## <tr><td style="text-align:left"></td><td>(0.034)</td><td>(1.417)</td><td>(0.045)</td><td>(0.036)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">No_postgrad</td><td>-0.014</td><td>-0.874</td><td>-0.021</td><td>0.00
## <tr><td style="text-align:left"></td><td>(0.021)</td><td>(1.115)</td><td>(0.028)</td><td>(0.022)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">No_professor</td><td>-0.017</td><td>0.469</td><td>-0.009</td><td>-0.0
## <tr><td style="text-align:left"></td><td>(0.032)</td><td>(1.290)</td><td>(0.042)</td><td>(0.033)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.universityCUHK</td><td>0.185<sup>**</sup></td><td>0.211</td><td>0.1
## <tr><td style="text-align:left"></td><td>(0.074)</td><td>(3.597)</td><td>(0.097)</td><td>(0.077)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.universityHKBU</td><td>0.019</td><td>-2.123</td><td>0.190</td><td>0
## <tr><td style="text-align:left"></td><td>(0.118)</td><td>(4.503)</td><td>(0.155)</td><td>(0.123)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.universityHKU</td><td>0.293<sup>***</sup></td><td>3.010</td><td>0.0
## <tr><td style="text-align:left"></td><td>(0.073)</td><td>(3.191)</td><td>(0.095)</td><td>(0.075)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.universityPoly</td><td>0.178<sup>**</sup></td><td>0.708</td><td>0.0
## <tr><td style="text-align:left"></td><td>(0.077)</td><td>(3.338)</td><td>(0.102)</td><td>(0.080)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.universityUST</td><td>0.151<sup>**</sup></td><td>1.374</td><td>-0.0
## <tr><td style="text-align:left"></td><td>(0.074)</td><td>(3.151)</td><td>(0.096)</td><td>(0.076)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2015</td><td>-0.017</td><td>8.211<sup>**</sup></td><td>0.133</t
## <tr><td style="text-align:left"></td><td>(0.090)</td><td>(3.576)</td><td>(0.118)</td><td>(0.093)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2016</td><td>0.005</td><td>0.344</td><td>0.032</td><td>0.070</t
## <tr><td style="text-align:left"></td><td>(0.082)</td><td>(2.766)</td><td>(0.108)</td><td>(0.086)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2017</td><td>0.140</td><td>0.082</td><td>0.073</td><td>-0.008</
## <tr><td style="text-align:left"></td><td>(0.089)</td><td>(3.317)</td><td>(0.116)</td><td>(0.092)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2018</td><td>0.059</td><td>-3.197</td><td>0.053</td><td>-0.009
## <tr><td style="text-align:left"></td><td>(0.087)</td><td>(3.069)</td><td>(0.115)</td><td>(0.091)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2019</td><td>0.056</td><td>-0.774</td><td>-0.014</td><td>-0.02
## <tr><td style="text-align:left"></td><td>(0.077)</td><td>(2.760)</td><td>(0.100)</td><td>(0.079)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.year2020</td><td>0.156<sup>*</sup></td><td>1.257</td><td>0.048</td
## <tr><td style="text-align:left"></td><td>(0.084)</td><td>(4.426)</td><td>(0.110)</td><td>(0.087)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaAutomotive Parts and Accessories</td><td>0.141</td><td></td><td
## <tr><td style="text-align:left"></td><td>(0.368)</td><td></td><td>(0.483)</td><td>(0.382)</td><td>(0
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaBiotechnology</td><td>0.047</td><td>-2.735</td><td>0.178</td><t
## <tr><td style="text-align:left"></td><td>(0.089)</td><td>(4.963)</td><td>(0.116)</td><td>(0.092)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaElectronics</td><td>0.123</td><td>-3.729</td><td>0.175</td><td
## <tr><td style="text-align:left"></td><td>(0.102)</td><td>(5.715)</td><td>(0.135)</td><td>(0.107)</td
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
```

```
## <tr><td style="text-align:left">f.areaEnergy</td><td>0.255</td><td></td><td>-0.119</td><td>0.077</td>
## <tr><td style="text-align:left"></td><td>(0.194)</td><td></td><td>(0.255)</td><td>(0.202)</td><td>(0
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaEnvironmental Protection</td><td>-0.171</td><td>-0.731</td><td
## <tr><td style="text-align:left"></td><td>(0.155)</td><td>(7.291)</td><td>(0.203)</td><td>(0.161)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaInformation and Communication Technologies</td><td>0.152<sup>*
## <tr><td style="text-align:left"></td><td>(0.086)</td><td>(4.759)</td><td>(0.112)</td><td>(0.089)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaLogistics and Supply Chain Management</td><td>0.186</td><td>-2
## <tr><td style="text-align:left"></td><td>(0.264)</td><td>(8.521)</td><td>(0.345)</td><td>(0.274)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaNanotechnology and Materials Science</td><td>0.189</td><td>0.3
## <tr><td style="text-align:left"></td><td>(0.123)</td><td>(6.811)</td><td>(0.161)</td><td>(0.127)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaOther</td><td>0.171</td><td>-0.457</td><td>0.430<sup>***</sup>
## <tr><td style="text-align:left"></td><td>(0.118)</td><td>(5.629)</td><td>(0.154)</td><td>(0.122)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaTesting and Certification</td><td>-0.124</td><td>-4.630</td><td
## <tr><td style="text-align:left"></td><td>(0.228)</td><td>(9.306)</td><td>(0.299)</td><td>(0.237)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">f.areaTextiles/Clothing/Footwear</td><td>-0.688<sup>*</sup></td><td
## <tr><td style="text-align:left"></td><td>(0.373)</td><td></td><td>(0.489)</td><td>(0.387)</td><td>(0
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">Constant</td><td>0.487<sup>***</sup></td><td>2.975</td><td>0.041</td>
## <tr><td style="text-align:left"></td><td>(0.123)</td><td>(6.018)</td><td>(0.162)</td><td>(0.128)</td>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
## <tr><td colspan="7" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.138</td><td>0.247</td><td>0.119</td><td>0.07
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.054</td><td>-0.023</td><td>0.033</td
## <tr><td style="text-align:left">Residual Std. Error</td><td>0.353 (df = 268)</td><td>7.013 (df = 64)
## <tr><td style="text-align:left">F Statistic</td><td>1.645<sup>**</sup> (df = 26; 268)</td><td>0.915
## <tr><td colspan="7" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## </table>
```

Question 3d

```
area = table(tsssu$area)
area = area[order(area,decreasing = FALSE)]
barplot(area,
        las=1,
        horiz=TRUE)
```


nologies
chnology
ectronics
elopment
Other
Science
rotection
Energy
tification
agement
Footwear
essories
0
20
40
60
80
100