Econometrics Data Assignment 1

Name: Utkarsh Arora

Roll No: 2020143

Q1)

First, we created datasets for gdp, beds and tap, using data from the internet from the links that were provided. Then we have added this data to the appended columns in the larger dataset.

We have performed data-cleaning as well. Our analysis in this assignment depends on the dependant variables v40, v42, v43, v44, v45, v46; and the explanatory variables index, gdp, beds, tap. For some states/districts, values for gdp and tap is not given (Eg Ladakh). For the years 2017-2019, we don't have data for the dependant variables. In any of the rows where data for any of the 10 variables (dependant and independent) is missing, we have removed that row. This is to make sure that our vectors when conducting analysis do not have NULL values and are of equal sizes.

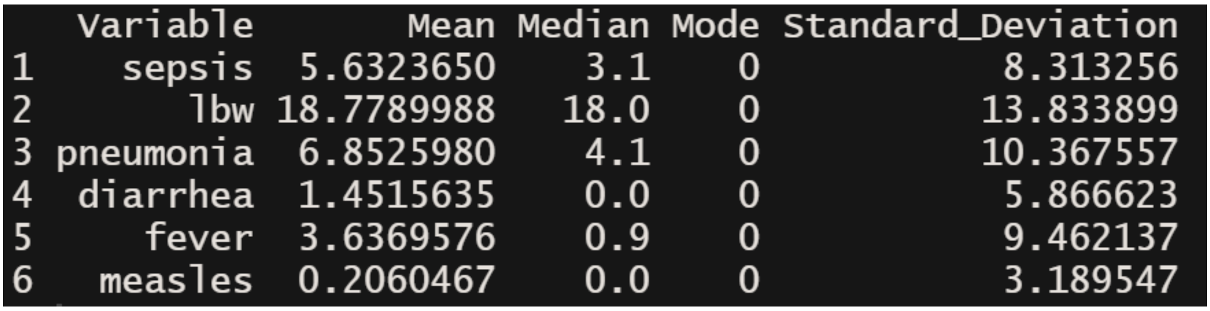
In other variables where we have missing data, we have put NA.

Input dataset: main.csv

Appended and cleaned dataset: main9.csv

Q2)

a) File: q2a.r

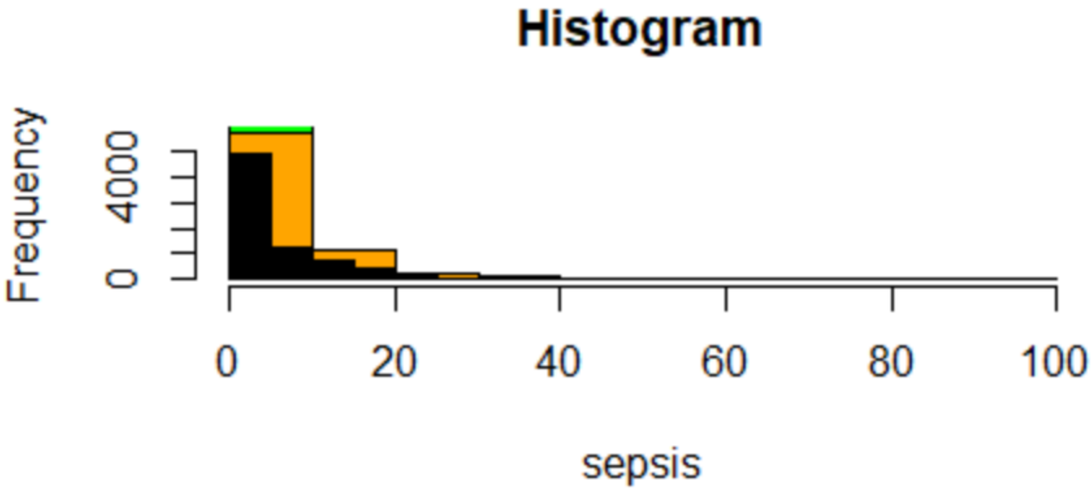


```
   Variable         Mean Median Mode Standard_Deviation
1    sepsis  5.6323650    3.1    0           8.313256
2       lbw 18.7789988   18.0    0          13.833899
3 pneumonia  6.8525980    4.1    0          10.367557
4  diarrhea  1.4515635    0.0    0           5.866623
5     fever  3.6369576    0.9    0           9.462137
6   measles  0.2060467    0.0    0           3.189547
```

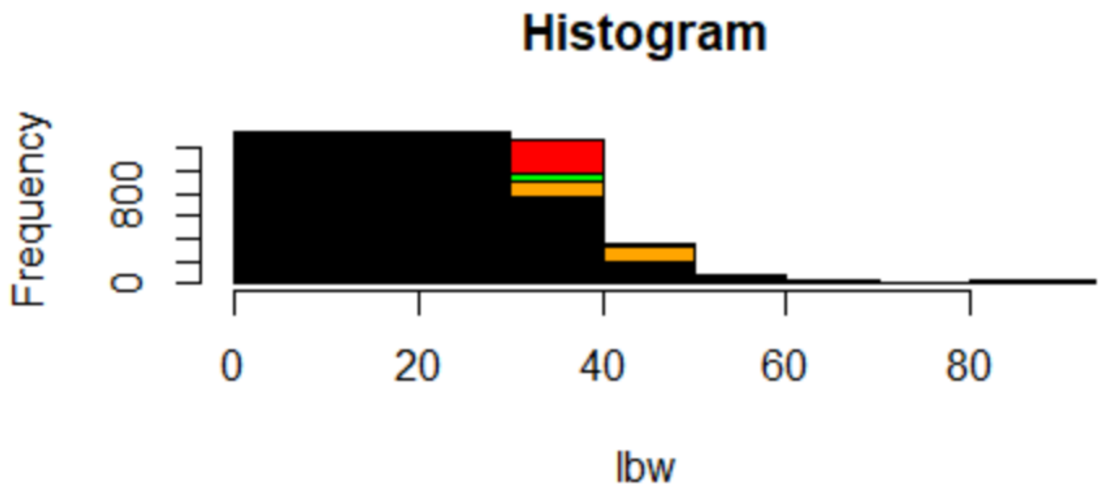
b)

YEAR-

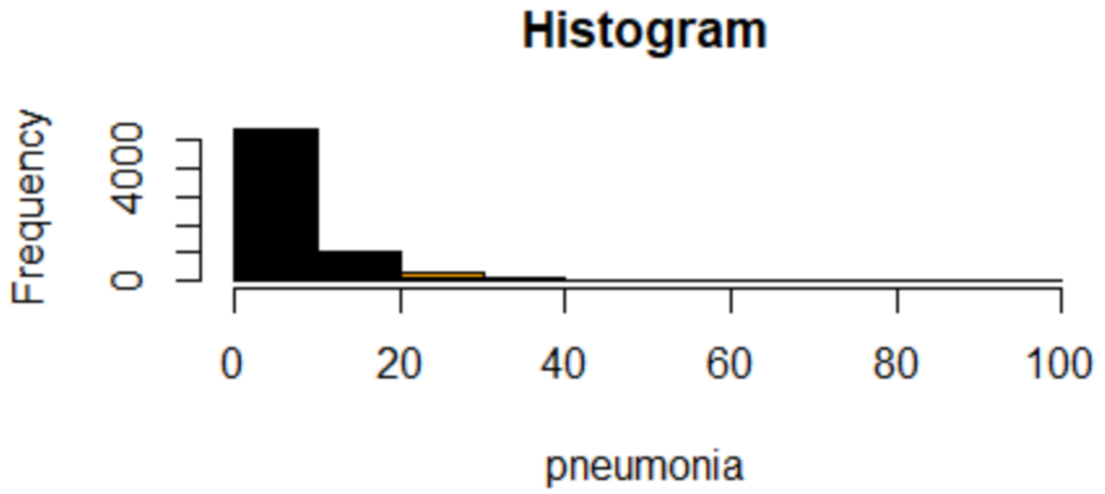
Sepsis



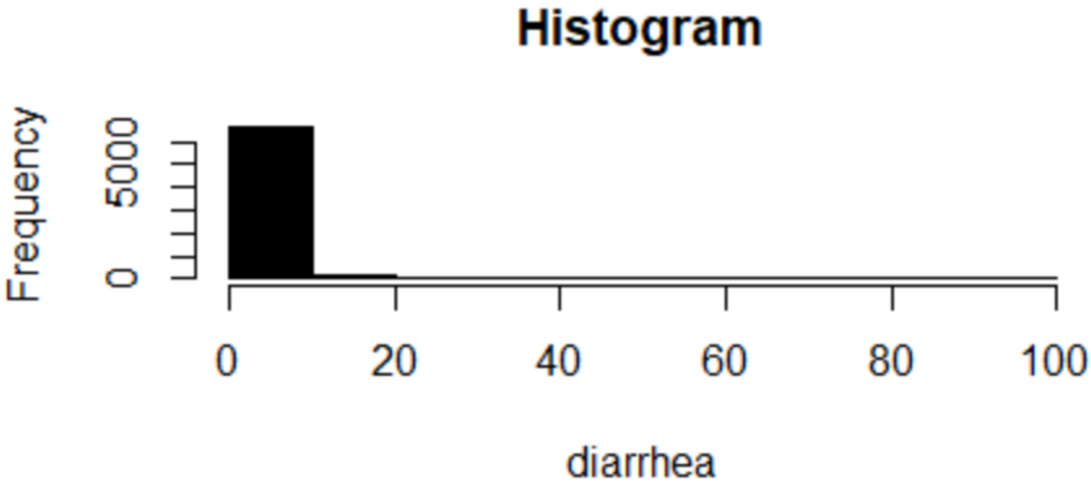
Lbw



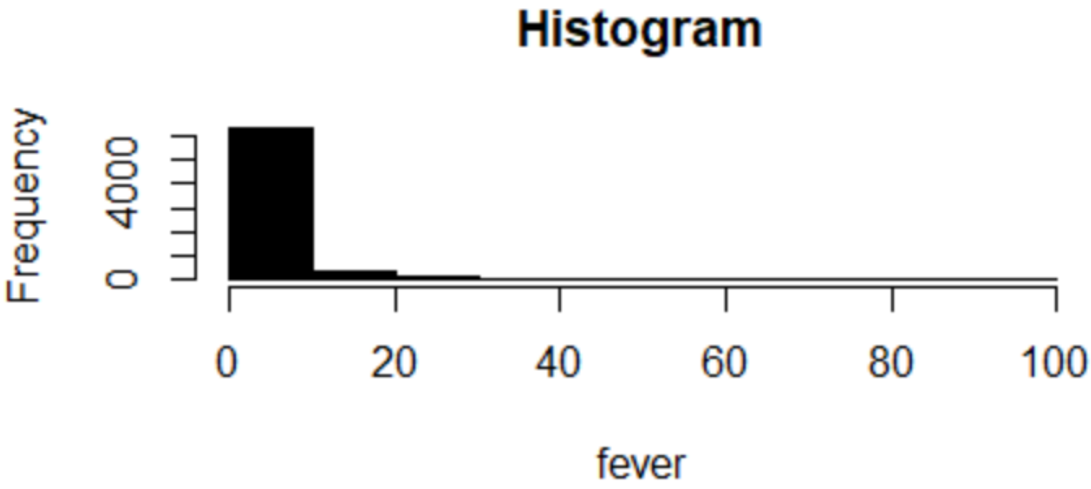
Pneumonia



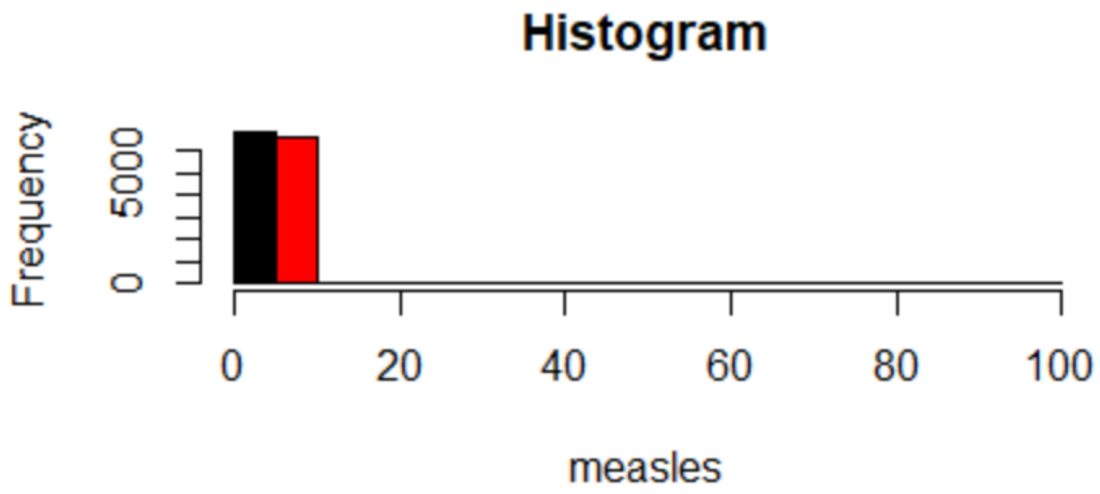
Diarrhoea



Fever

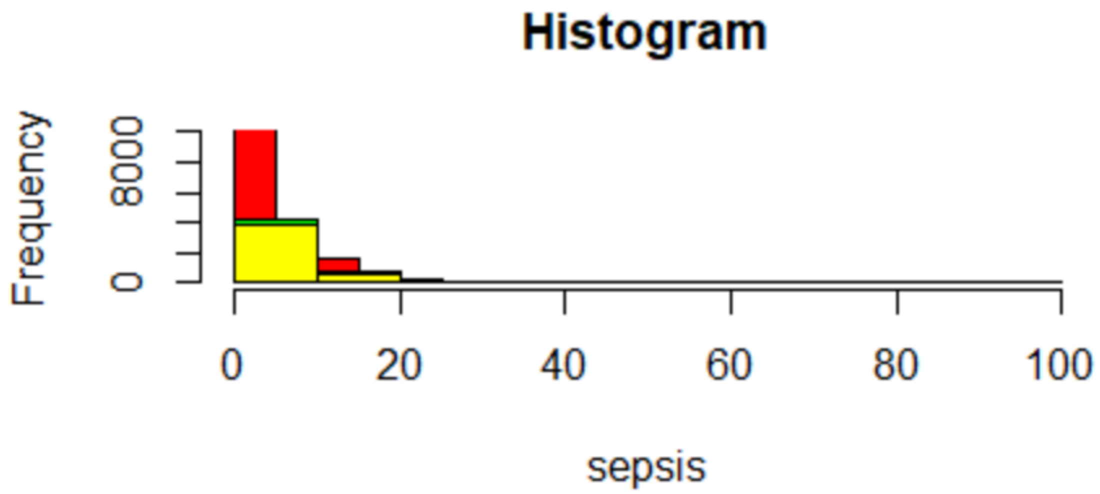


Measles

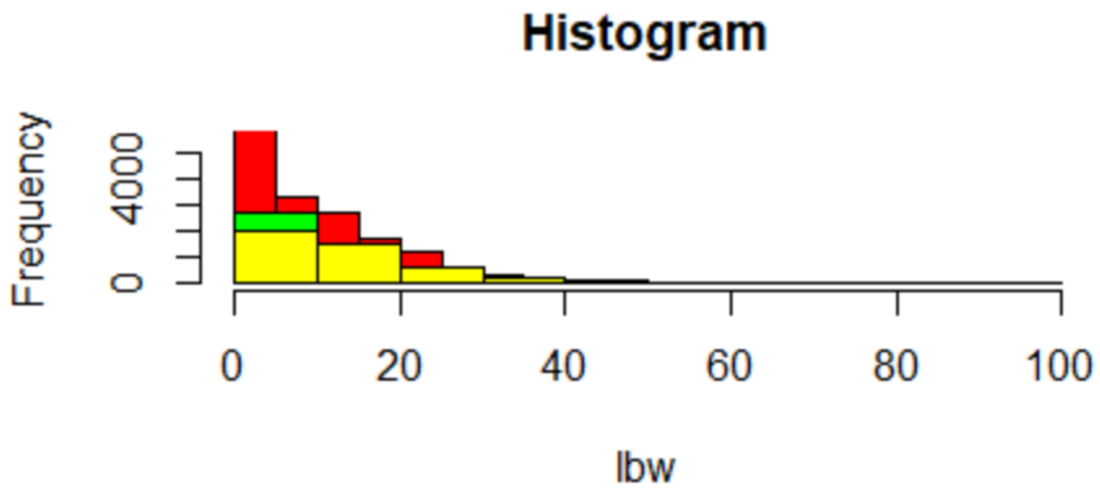


Season--

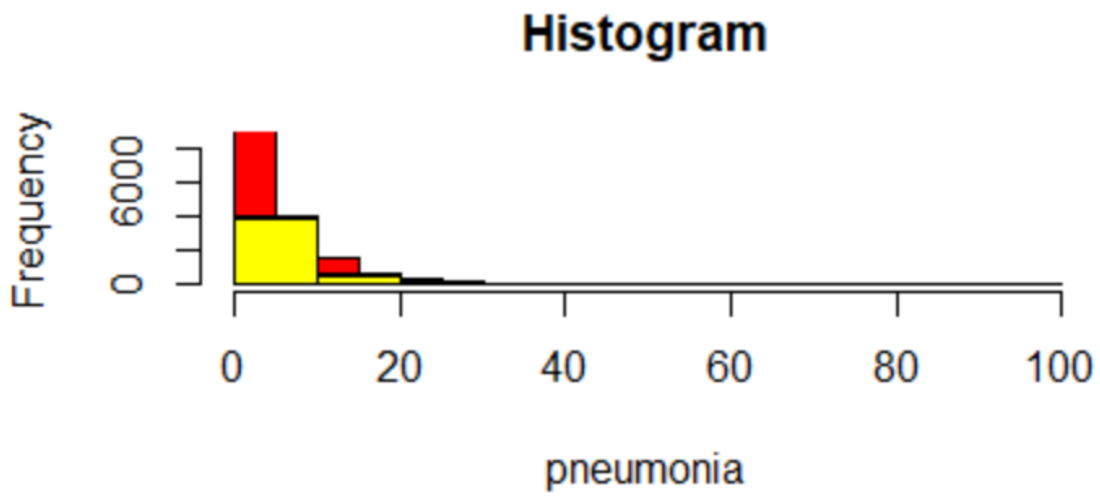
Sepsis



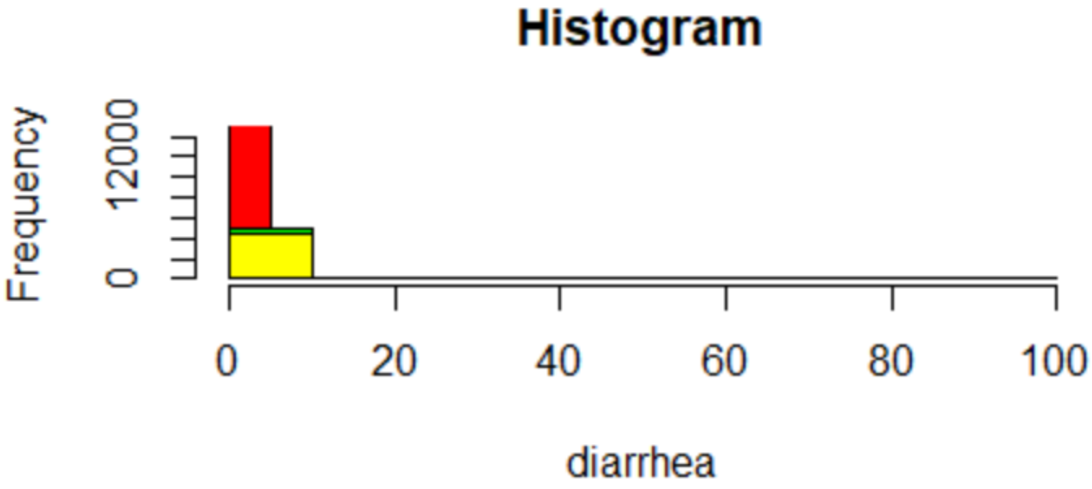
Lbw



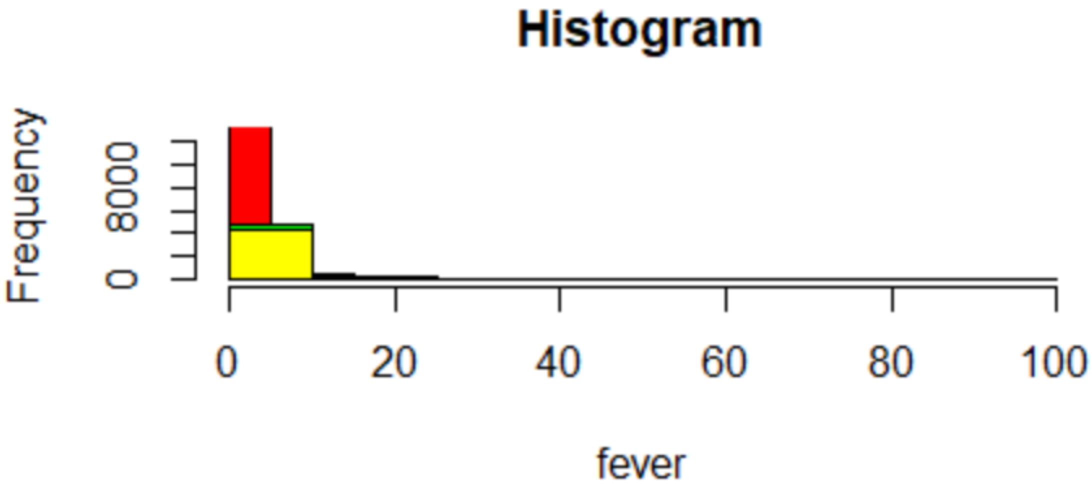
Pneumonia



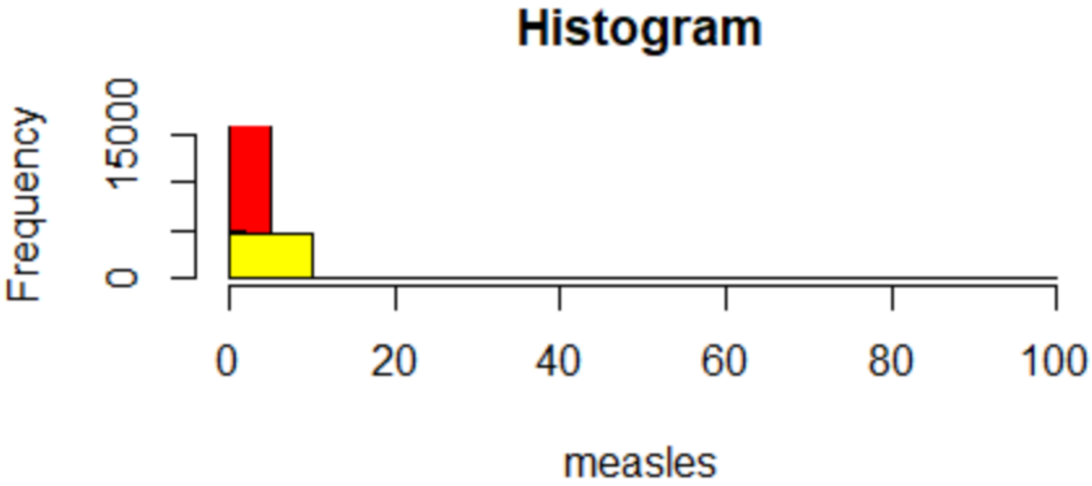
Diahhroea



Fever

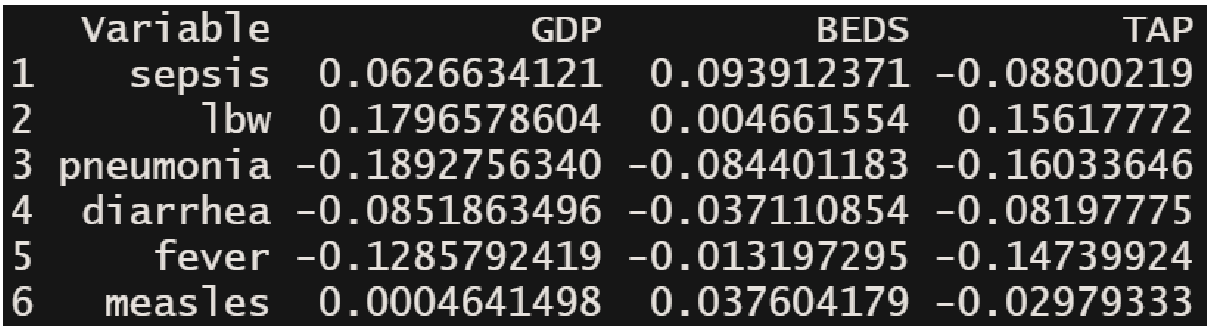


Measles



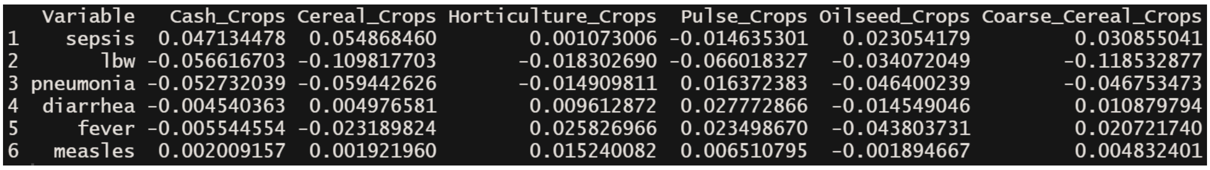
d)

1) File: q2d1.r



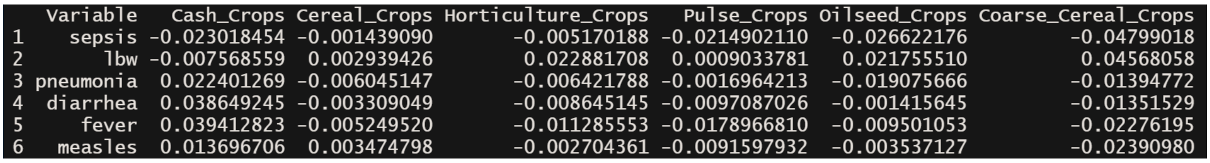
```
   Variable           GDP        BEDS         TAP
1    sepsis  0.0626634121  0.093912371 -0.08800219
2       lbw  0.1796578604  0.004661554  0.15617772
3 pneumonia -0.1892756340 -0.084401183 -0.16033646
4  diarrhea -0.0851863496 -0.037110854 -0.08197775
5     fever -0.1285792419 -0.013197295 -0.14739924
6   measles  0.0004641498  0.037604179 -0.02979333
```

2) File: q2d2.r



```
   Variable   Cash_Crops Cereal_Crops Horticulture_Crops  Pulse_Crops Oilseed_Crops Coarse_Cereal_Crops
1    sepsis  0.047134478  0.054868460        0.001073006 -0.014635301   0.023054179         0.030855041
2       lbw -0.056616703 -0.109817703       -0.018302690 -0.066018327  -0.034072049        -0.118532877
3 pneumonia -0.052732039 -0.059442626       -0.014909811  0.016372383  -0.046400239        -0.046753473
4  diarrhea -0.004540363  0.004976581        0.009612872  0.027772866  -0.014549046         0.010879794
5     fever -0.005544554 -0.023189824        0.025826966  0.023498670  -0.043803731         0.020721740
6   measles  0.002009157  0.001921960        0.015240082  0.006510795  -0.001894667         0.004832401
```

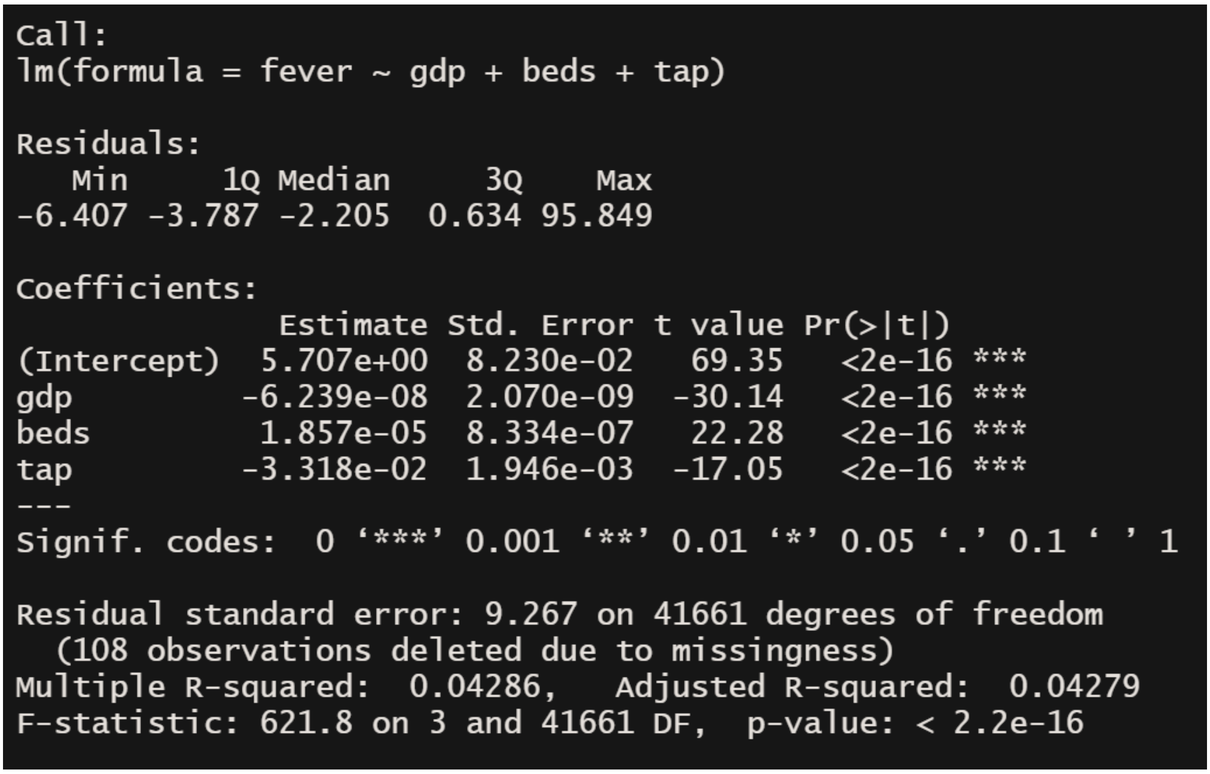
3) File: q2d3.r



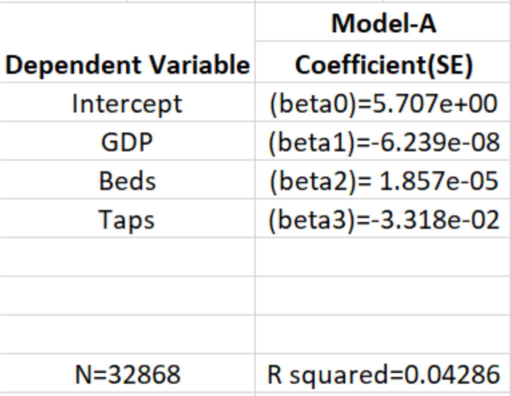
```
   Variable   Cash_Crops Cereal_Crops Horticulture_Crops   Pulse_Crops Oilseed_Crops Coarse_Cereal_Crops
1    sepsis -0.023018454 -0.001439090       -0.005170188 -0.0214902110  -0.026622176         -0.04799018
2       lbw -0.007568559  0.002939426        0.022881708  0.0009033781   0.021755510          0.04568058
3 pneumonia  0.022401269 -0.006045147       -0.006421788 -0.0016964213  -0.019075666         -0.01394772
4  diarrhea  0.038649245 -0.003309049       -0.008645145 -0.0097087026  -0.001415645         -0.01351529
5     fever  0.039412823 -0.005249520       -0.011285553 -0.0178966810  -0.009501053         -0.02276195
6   measles  0.013696706  0.003474798       -0.002704361 -0.0091597932  -0.003537127         -0.02390980
```

Q3)

A) File: q3a.r

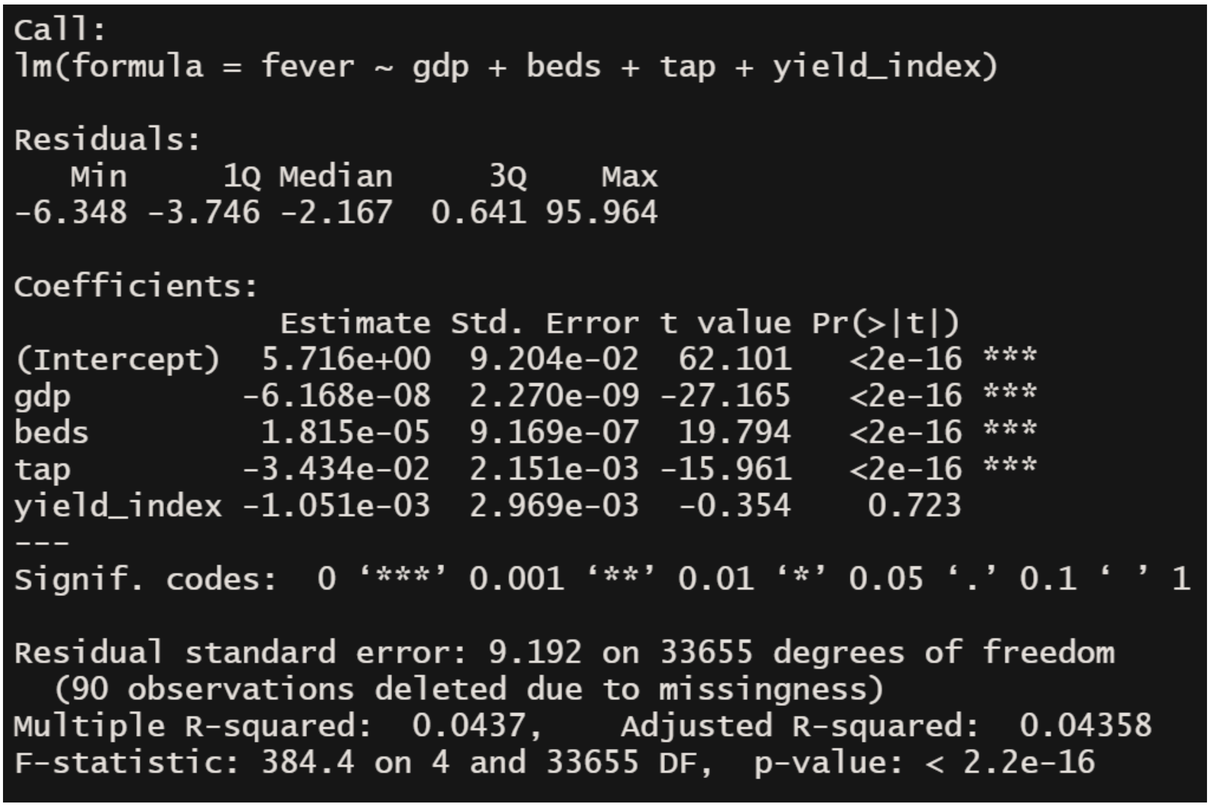


```
Call:
lm(formula = fever ~ gdp + beds + tap)

Residuals:
   Min     1Q Median     3Q    Max 
-6.407 -3.787 -2.205  0.634 95.849 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.707e+00  8.230e-02   69.35   <2e-16 ***
gdp         -6.239e-08  2.070e-09  -30.14   <2e-16 ***
beds         1.857e-05  8.334e-07   22.28   <2e-16 ***
tap         -3.318e-02  1.946e-03  -17.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.267 on 41661 degrees of freedom
  (108 observations deleted due to missingness)
Multiple R-squared:  0.04286,	Adjusted R-squared:  0.04279 
F-statistic: 621.8 on 3 and 41661 DF,  p-value: < 2.2e-16
```

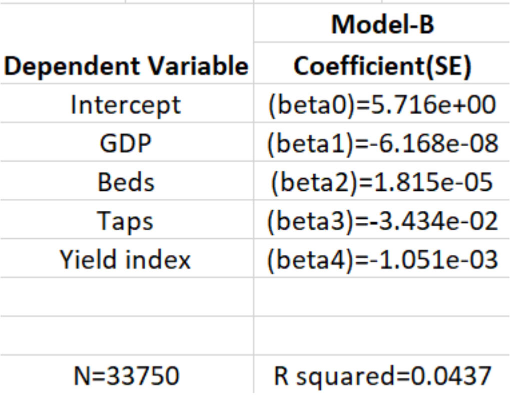


	Model-A
Dependent Variable	Coefficient(SE)
Intercept	(beta0)=5.707e+00
GDP	(beta1)=-6.239e-08
Beds	(beta2)= 1.857e-05
Taps	(beta3)=-3.318e-02
N=32868	R squared=0.04286

B) File: q3b.r

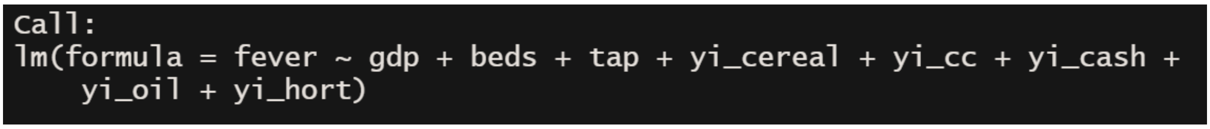


```
Call:
lm(formula = fever ~ gdp + beds + tap + yield_index)

Residuals:
   Min     1Q Median     3Q    Max 
-6.348 -3.746 -2.167  0.641 95.964 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.716e+00  9.204e-02  62.101   <2e-16 ***
gdp         -6.168e-08  2.270e-09 -27.165   <2e-16 ***
beds         1.815e-05  9.169e-07  19.794   <2e-16 ***
tap         -3.434e-02  2.151e-03 -15.961   <2e-16 ***
yield_index -1.051e-03  2.969e-03  -0.354    0.723    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.192 on 33655 degrees of freedom
  (90 observations deleted due to missingness)
Multiple R-squared:  0.0437,	Adjusted R-squared:  0.04358 
F-statistic: 384.4 on 4 and 33655 DF,  p-value: < 2.2e-16
```

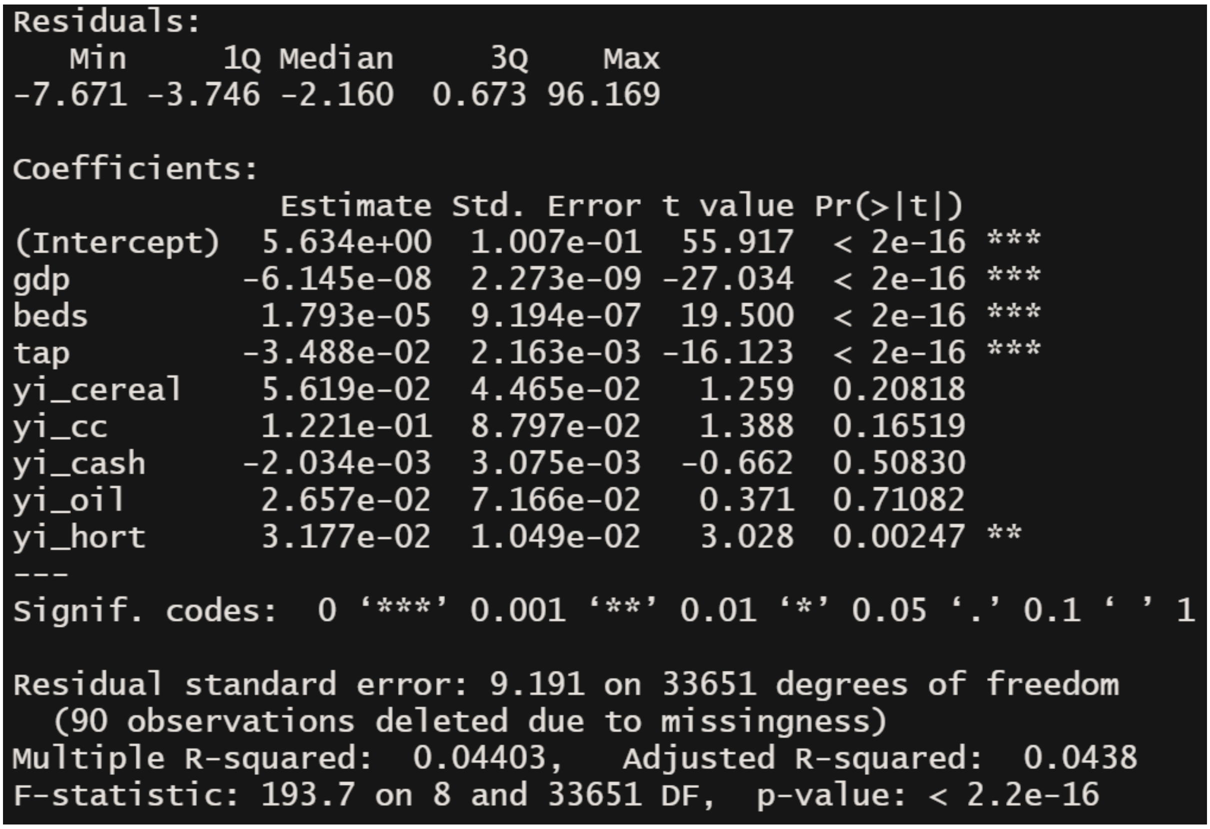


	Model-B
Dependent Variable	**Coefficient(SE)**
Intercept	(beta0)=5.716e+00
GDP	(beta1)=-6.168e-08
Beds	(beta2)=1.815e-05
Taps	(beta3)=-3.434e-02
Yield index	(beta4)=-1.051e-03
N=33750	R squared=0.0437

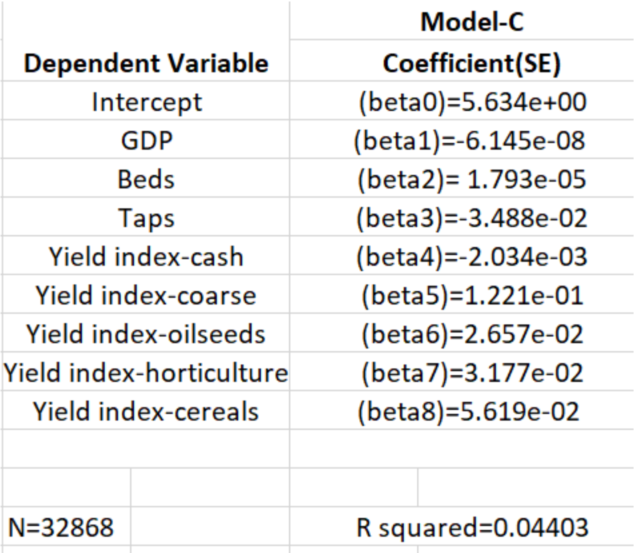
C) File: q3c.r



```
Call:
lm(formula = fever ~ gdp + beds + tap + yi_cereal + yi_cc + yi_cash + 
    yi_oil + yi_hort)
```

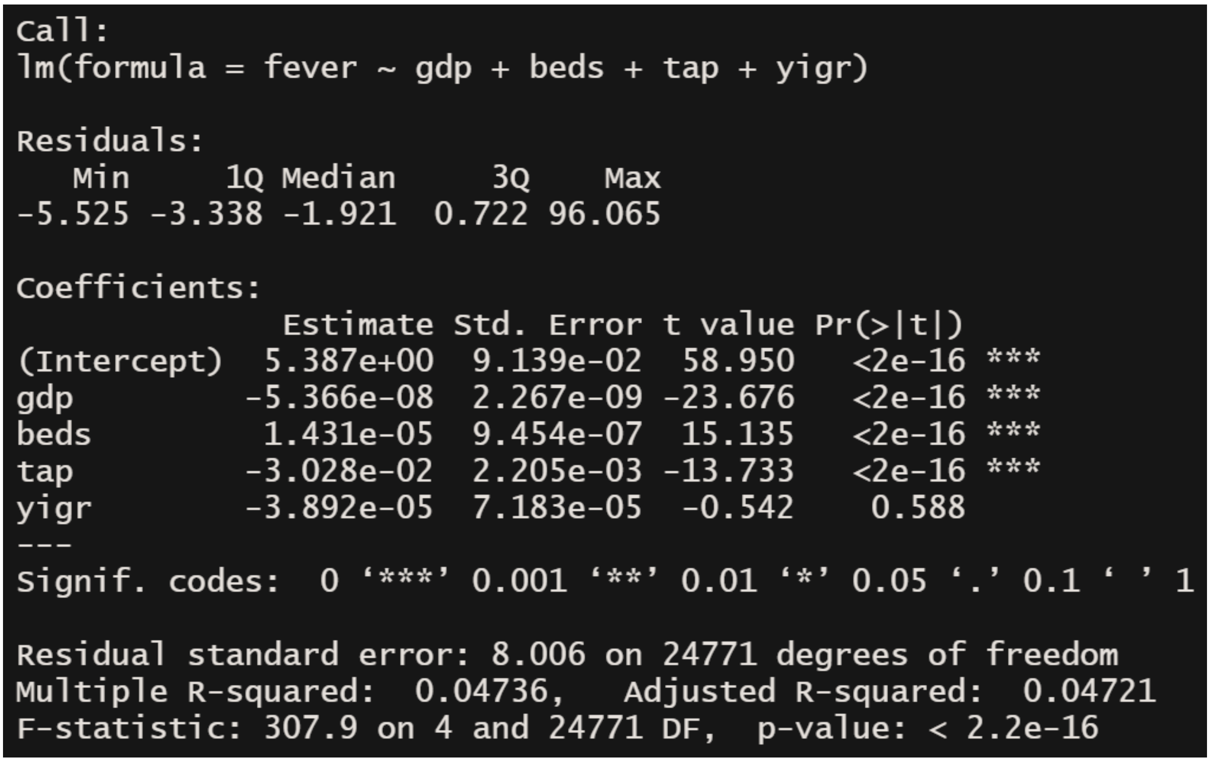


```
Residuals:
   Min     1Q Median     3Q    Max
-7.671 -3.746 -2.160  0.673 96.169

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.634e+00  1.007e-01  55.917  < 2e-16 ***
gdp         -6.145e-08  2.273e-09 -27.034  < 2e-16 ***
beds         1.793e-05  9.194e-07  19.500  < 2e-16 ***
tap         -3.488e-02  2.163e-03 -16.123  < 2e-16 ***
yi_cereal    5.619e-02  4.465e-02   1.259  0.20818
yi_cc        1.221e-01  8.797e-02   1.388  0.16519
yi_cash     -2.034e-03  3.075e-03  -0.662  0.50830
yi_oil       2.657e-02  7.166e-02   0.371  0.71082
yi_hort      3.177e-02  1.049e-02   3.028  0.00247 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.191 on 33651 degrees of freedom
  (90 observations deleted due to missingness)
Multiple R-squared:  0.04403,	Adjusted R-squared:  0.0438
F-statistic: 193.7 on 8 and 33651 DF,  p-value: < 2.2e-16
```

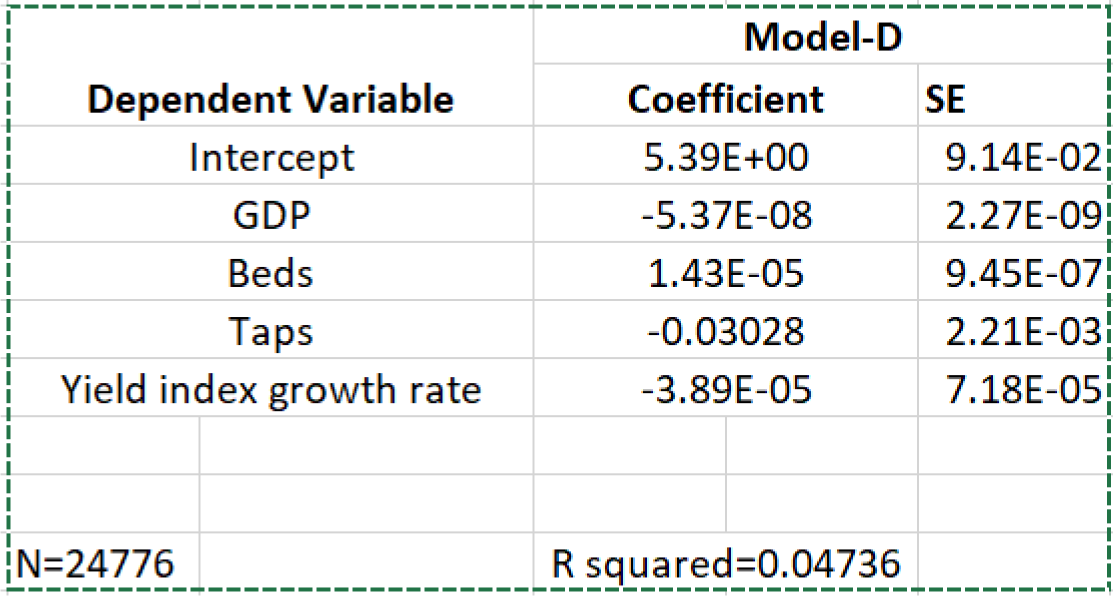


	Model-C
Dependent Variable	**Coefficient(SE)**
Intercept	(beta0)=5.634e+00
GDP	(beta1)=-6.145e-08
Beds	(beta2)= 1.793e-05
Taps	(beta3)=-3.488e-02
Yield index-cash	(beta4)=-2.034e-03
Yield index-coarse	(beta5)=1.221e-01
Yield index-oilseeds	(beta6)=2.657e-02
Yield index-horticulture	(beta7)=3.177e-02
Yield index-cereals	(beta8)=5.619e-02
N=32868	R squared=0.04403

D) File: q3d.r

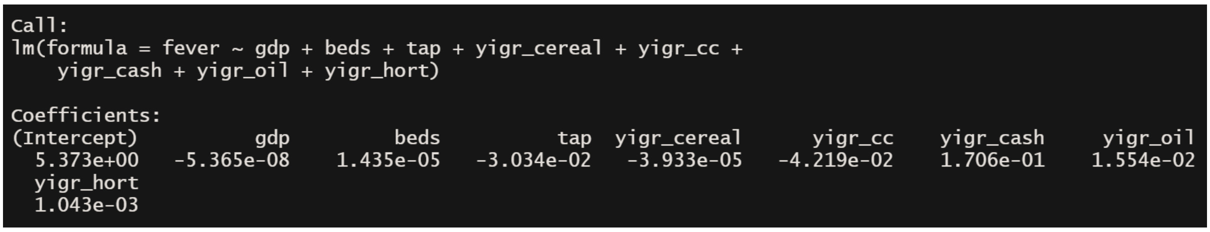


```
Call:
lm(formula = fever ~ gdp + beds + tap + yigr)

Residuals:
   Min     1Q Median     3Q    Max 
-5.525 -3.338 -1.921  0.722 96.065 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.387e+00  9.139e-02  58.950   <2e-16 ***
gdp         -5.366e-08  2.267e-09 -23.676   <2e-16 ***
beds         1.431e-05  9.454e-07  15.135   <2e-16 ***
tap         -3.028e-02  2.205e-03 -13.733   <2e-16 ***
yigr        -3.892e-05  7.183e-05  -0.542    0.588    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.006 on 24771 degrees of freedom
Multiple R-squared:  0.04736,	Adjusted R-squared:  0.04721 
F-statistic: 307.9 on 4 and 24771 DF,  p-value: < 2.2e-16
```

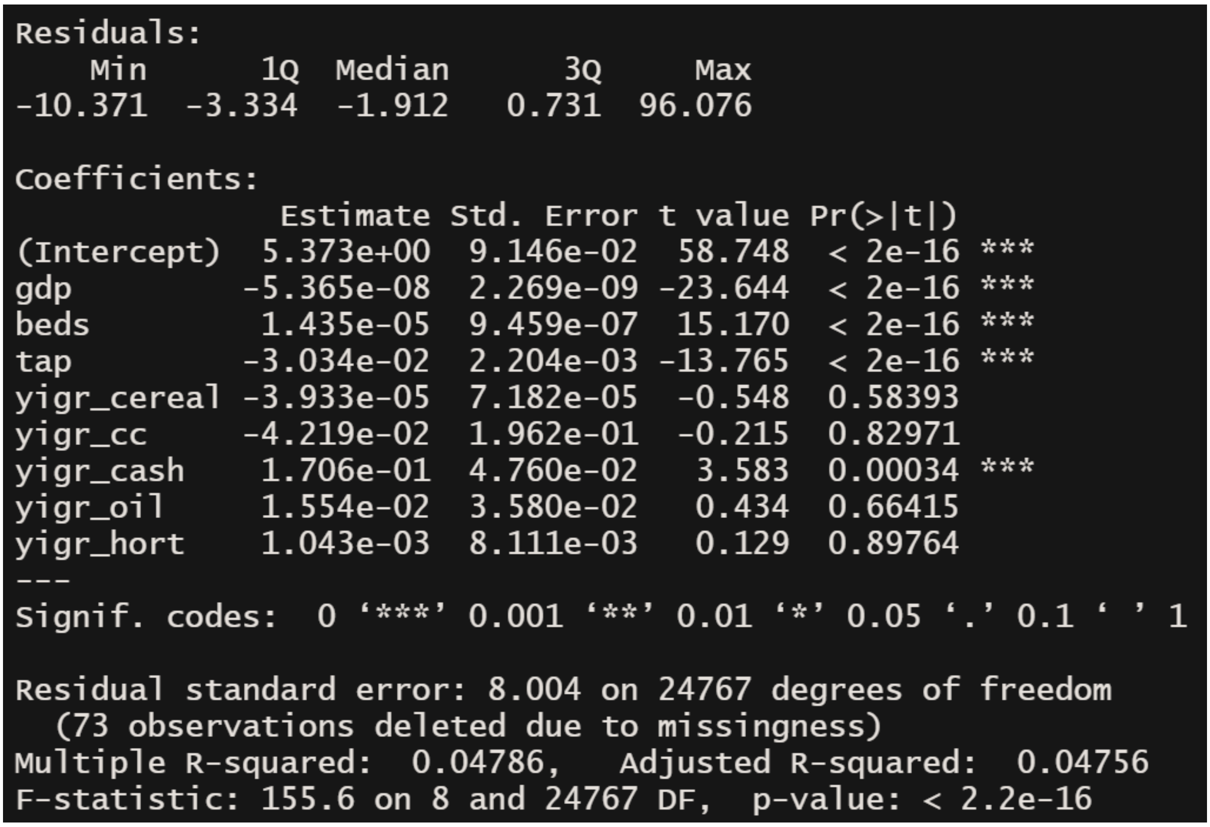


	Model-D	
Dependent Variable	**Coefficient**	**SE**
Intercept	5.39E+00	9.14E-02
GDP	-5.37E-08	2.27E-09
Beds	1.43E-05	9.45E-07
Taps	-0.03028	2.21E-03
Yield index growth rate	-3.89E-05	7.18E-05
N=24776	R squared=0.04736	

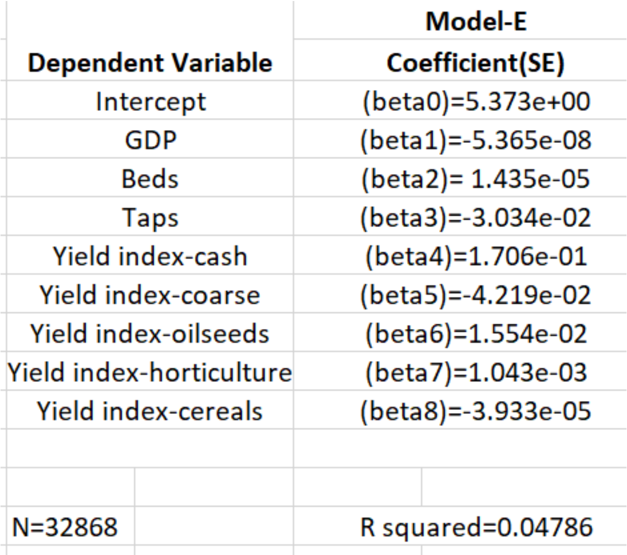
E) File: q3e.r



```
Call:
lm(formula = fever ~ gdp + beds + tap + yigr_cereal + yigr_cc + 
    yigr_cash + yigr_oil + yigr_hort)

Coefficients:
(Intercept)          gdp         beds          tap  yigr_cereal      yigr_cc    yigr_cash     yigr_oil  
  5.373e+00   -5.365e-08    1.435e-05   -3.034e-02   -3.933e-05   -4.219e-02    1.706e-01    1.554e-02  
  yigr_hort  
  1.043e-03  
```

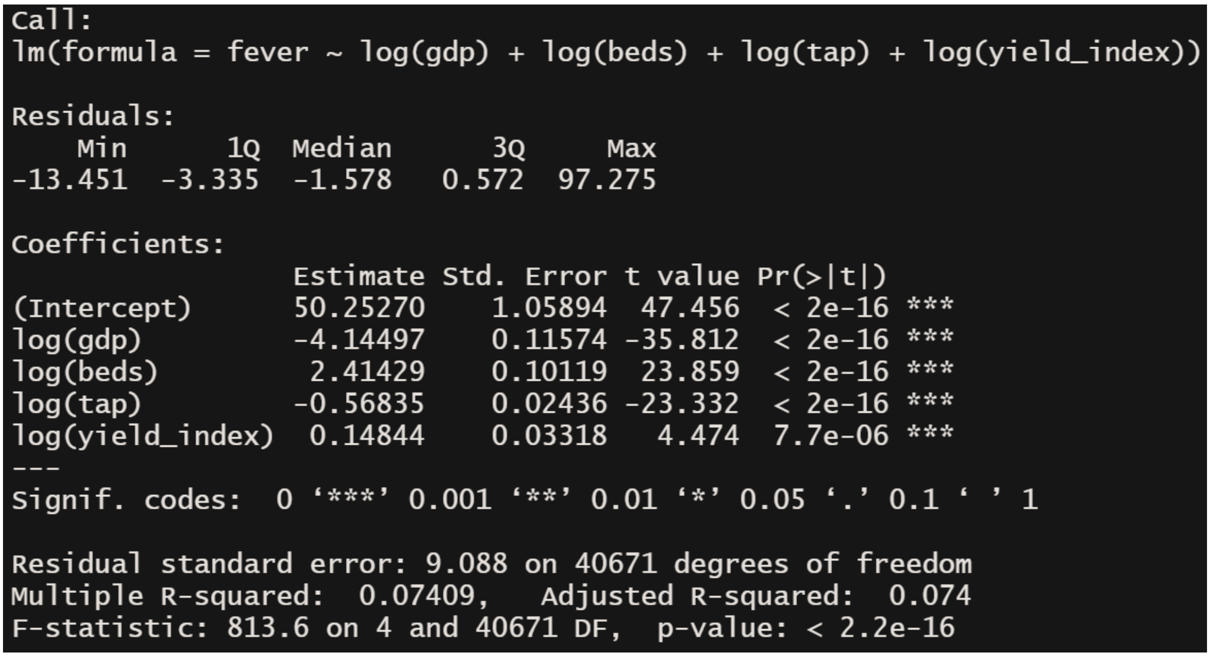


```
Residuals:
    Min      1Q  Median      3Q     Max 
-10.371  -3.334  -1.912   0.731  96.076 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.373e+00  9.146e-02  58.748  < 2e-16 ***
gdp         -5.365e-08  2.269e-09 -23.644  < 2e-16 ***
beds         1.435e-05  9.459e-07  15.170  < 2e-16 ***
tap         -3.034e-02  2.204e-03 -13.765  < 2e-16 ***
yigr_cereal -3.933e-05  7.182e-05  -0.548  0.58393    
yigr_cc     -4.219e-02  1.962e-01  -0.215  0.82971    
yigr_cash    1.706e-01  4.760e-02   3.583  0.00034 ***
yigr_oil     1.554e-02  3.580e-02   0.434  0.66415    
yigr_hort    1.043e-03  8.111e-03   0.129  0.89764    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.004 on 24767 degrees of freedom
  (73 observations deleted due to missingness)
Multiple R-squared:  0.04786,	Adjusted R-squared:  0.04756 
F-statistic: 155.6 on 8 and 24767 DF,  p-value: < 2.2e-16
```

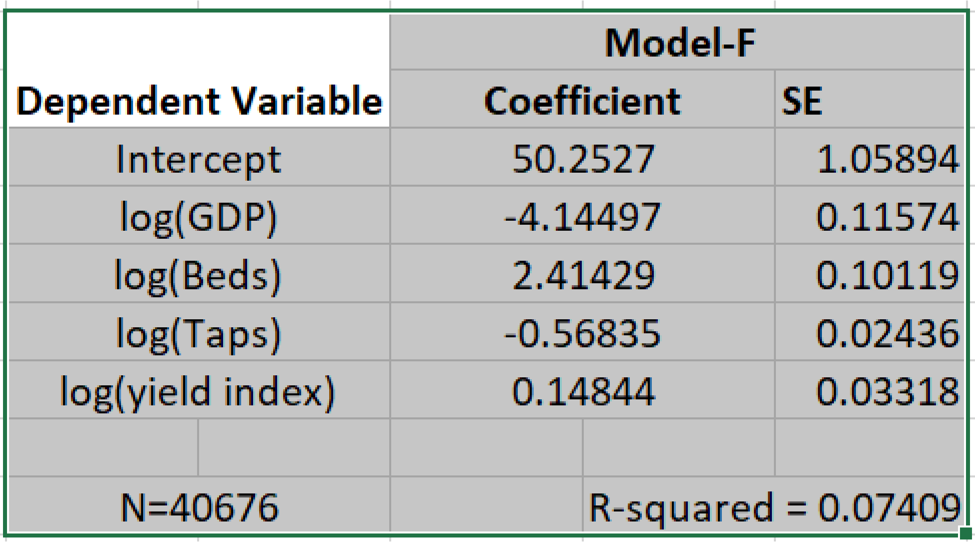


	Model-E
Dependent Variable	**Coefficient(SE)**
Intercept	(beta0)=5.373e+00
GDP	(beta1)=-5.365e-08
Beds	(beta2)= 1.435e-05
Taps	(beta3)=-3.034e-02
Yield index-cash	(beta4)=1.706e-01
Yield index-coarse	(beta5)=-4.219e-02
Yield index-oilseeds	(beta6)=1.554e-02
Yield index-horticulture	(beta7)=1.043e-03
Yield index-cereals	(beta8)=-3.933e-05
N=32868	R squared=0.04786

F) File: q3f.r

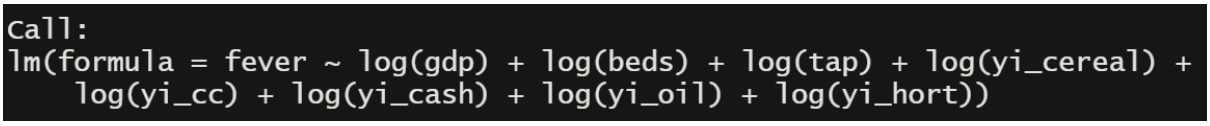


```
Call:
lm(formula = fever ~ log(gdp) + log(beds) + log(tap) + log(yield_index))

Residuals:
    Min      1Q  Median      3Q     Max 
-13.451  -3.335  -1.578   0.572  97.275 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      50.25270    1.05894  47.456  < 2e-16 ***
log(gdp)         -4.14497    0.11574 -35.812  < 2e-16 ***
log(beds)         2.41429    0.10119  23.859  < 2e-16 ***
log(tap)         -0.56835    0.02436 -23.332  < 2e-16 ***
log(yield_index)  0.14844    0.03318   4.474  7.7e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.088 on 40671 degrees of freedom
Multiple R-squared:  0.07409,	Adjusted R-squared:  0.074 
F-statistic: 813.6 on 4 and 40671 DF,  p-value: < 2.2e-16
```

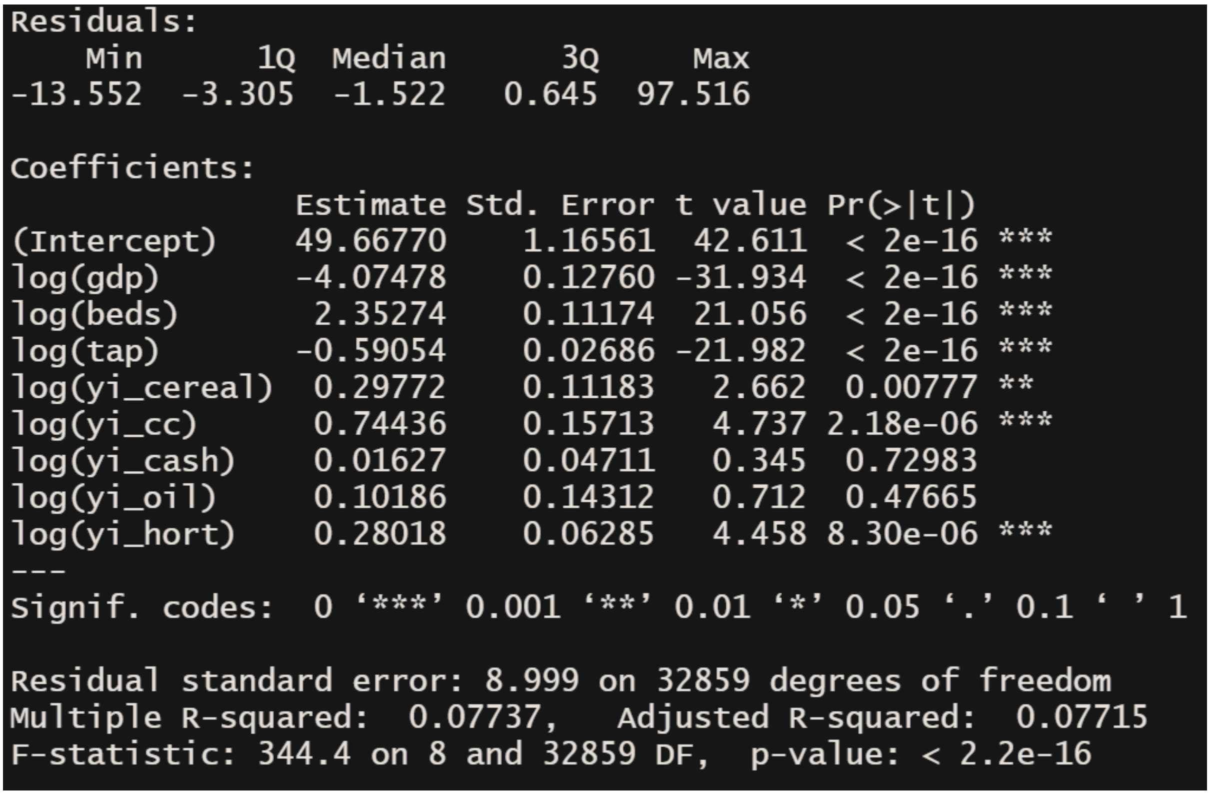


	Model-F	
Dependent Variable	**Coefficient**	**SE**
Intercept	50.2527	1.05894
log(GDP)	-4.14497	0.11574
log(Beds)	2.41429	0.10119
log(Taps)	-0.56835	0.02436
log(yield index)	0.14844	0.03318
N=40676		R-squared = 0.07409

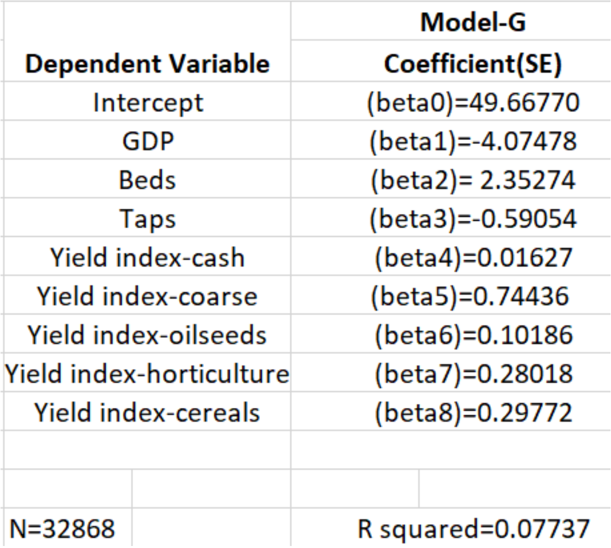
G) File: q3g.r



```
Call:
lm(formula = fever ~ log(gdp) + log(beds) + log(tap) + log(yi_cereal) + 
    log(yi_cc) + log(yi_cash) + log(yi_oil) + log(yi_hort))
```



```
Residuals:
    Min      1Q  Median      3Q     Max 
-13.552  -3.305  -1.522   0.645  97.516 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    49.66770    1.16561  42.611  < 2e-16 ***
log(gdp)       -4.07478    0.12760 -31.934  < 2e-16 ***
log(beds)       2.35274    0.11174  21.056  < 2e-16 ***
log(tap)       -0.59054    0.02686 -21.982  < 2e-16 ***
log(yi_cereal)  0.29772    0.11183   2.662  0.00777 ** 
log(yi_cc)      0.74436    0.15713   4.737 2.18e-06 ***
log(yi_cash)    0.01627    0.04711   0.345  0.72983    
log(yi_oil)     0.10186    0.14312   0.712  0.47665    
log(yi_hort)    0.28018    0.06285   4.458 8.30e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.999 on 32859 degrees of freedom
Multiple R-squared:  0.07737,	Adjusted R-squared:  0.07715 
F-statistic: 344.4 on 8 and 32859 DF,  p-value: < 2.2e-16
```



	Model-G
Dependent Variable	**Coefficient(SE)**
Intercept	(beta0)=49.66770
GDP	(beta1)=-4.07478
Beds	(beta2)= 2.35274
Taps	(beta3)=-0.59054
Yield index-cash	(beta4)=0.01627
Yield index-coarse	(beta5)=0.74436
Yield index-oilseeds	(beta6)=0.10186
Yield index-horticulture	(beta7)=0.28018
Yield index-cereals	(beta8)=0.29772
N=32868	R squared=0.07737

Q4) After running multiple analysis on the coefficient of correlation and R-Squared, we can conclude that in the given data, the theoretical relation between the coefficient of correlation and the goodness of fit does infact hold.

For a smaller example to display it better, correlation coefficient between fever and gdp is -0.1313496, and

Goodness of Fit when fever is regressed on gdp is 0.01725; which is exactly the square of the correlation.

Q5)

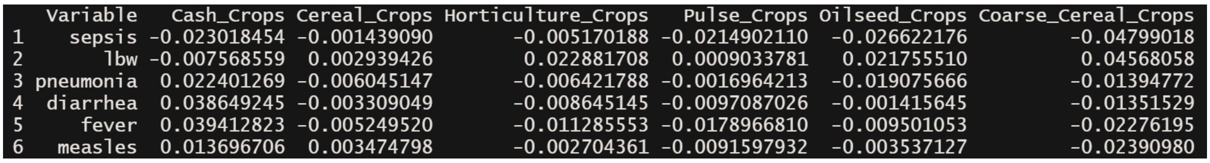
In the regression models we have analysed in which we take yield index of different crop categories separately, in many of them we can notice the fact:

If we keep all other explanatory variables constant, and just move yield index of cash crops, this will give an effect in the opposite direction, than if we kept all other explanatory variables constant and just moved yield index of coarse cereals.

Hence yield indexes of some of different crop categories have an opposing effect on the health indicator. This nuance is cancelled out and missing when we include the yield index for all six crop categories together. Thus we would be missing precision in exchange for generality.

Q6)

We obtained coefficients of correlation between yield growth and health indicators across crop categories in Q2D3



```
   Variable   Cash_Crops Cereal_Crops Horticulture_Crops   Pulse_Crops Oilseed_Crops Coarse_Cereal_Crops
1    sepsis -0.023018454 -0.001439090       -0.005170188 -0.0214902110  -0.026622176         -0.04799018
2       lbw -0.007568559  0.002939426        0.022881708  0.0009033781   0.021755510          0.04568058
3 pneumonia  0.022401269 -0.006045147       -0.006421788 -0.0016964213  -0.019075666         -0.01394772
4  diarrhea  0.038649245 -0.003309049       -0.008645145 -0.0097087026  -0.001415645         -0.01351529
5     fever  0.039412823 -0.005249520       -0.011285553 -0.0178966810  -0.009501053         -0.02276195
6   measles  0.013696706  0.003474798       -0.002704361 -0.0091597932  -0.003537127         -0.02390980
```

Sepsis has a negative correlation with yield growth across all crop categories.

lbw has a negative correlation with yield growth rate of cash crops, but a positive correlation with yield growth rates across other crop categories.

pneumonia has a positive correlation with yield growth rate of cash, but a negative correlation with yield growth rates across other crop categories.

Looking at this chart, we can conclude that the relation between yield growth and health indicators is not similar across crop categories.