**Econometrics Data Assignment**

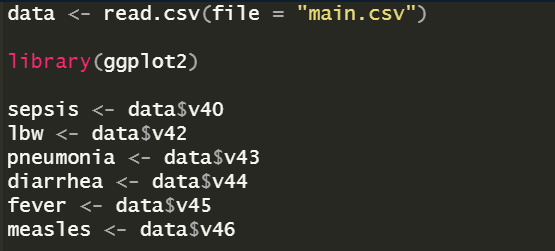
**Q1.**

Added Following data:

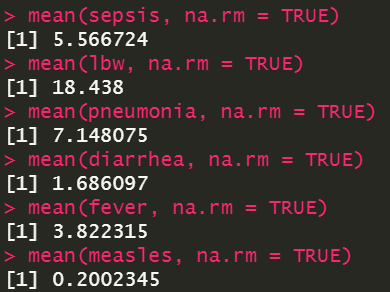
1. State Wise GDP Yearly gdp
2. State Wise Number of Hospital Beds (as of 2020)
3. District Wise Tap Water Access (Percentage of Households) as of 2019

In main.csv

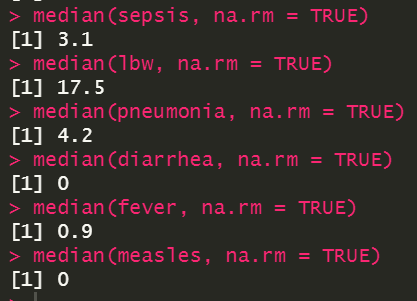
**Q2.**

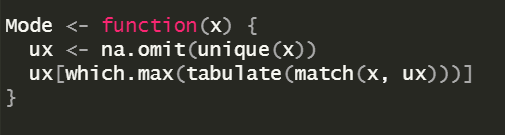
A.

Mean Data

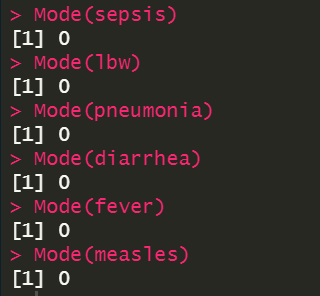


Median Data

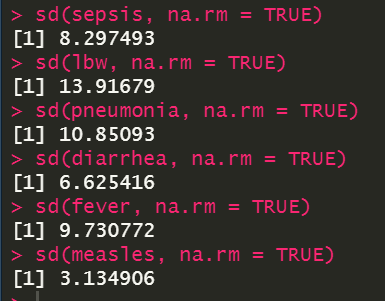




Mode Data



Standard Deviation Data

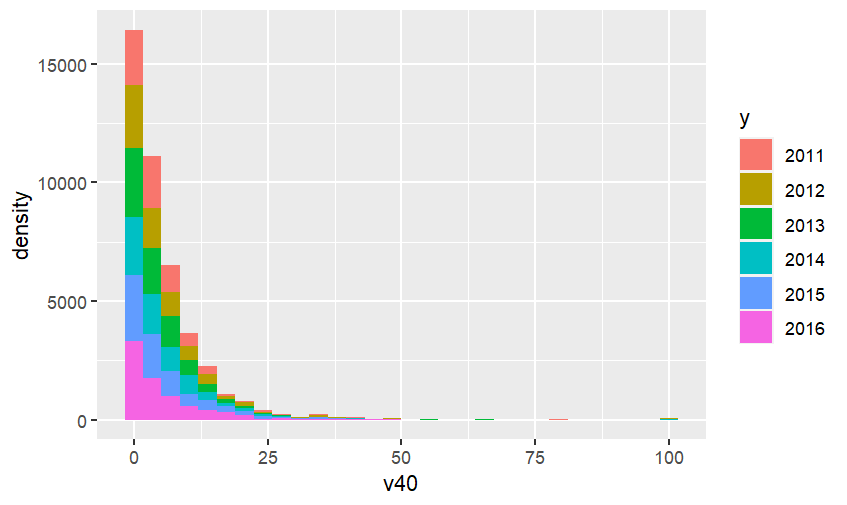


B.

Year-wise Histogram for percentage of infants death from sepsis.  
x axis -> percentage of infants death from sepsis  
y axis -> frequency of percentage of infants death from sepsis lying in the bin regions

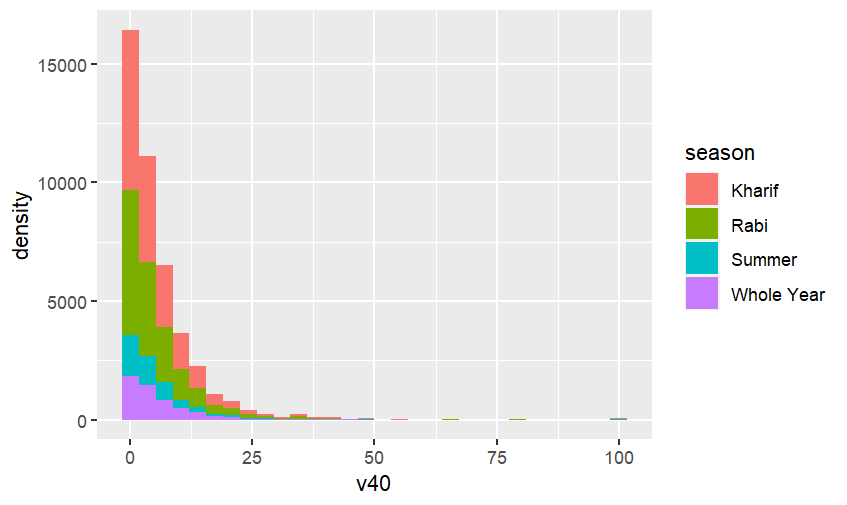
Colored graph -> Data for different years with the index on the right

Observation: The frequency %age has decreased over the years.

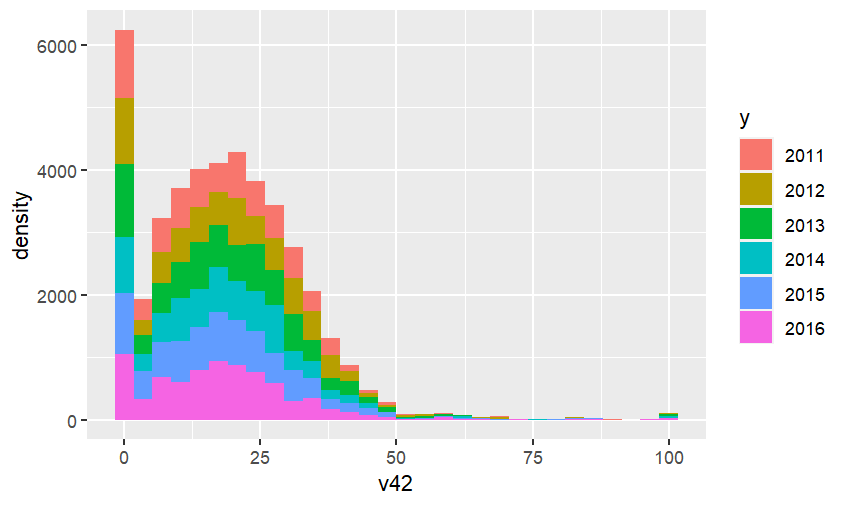


Season-wise Histogram for the percentage of infants death from sepsis.  
x axis -> percentage of infants death from sepsis  
y axis -> frequency of percentage of infants death from sepsis lying in the bin regions

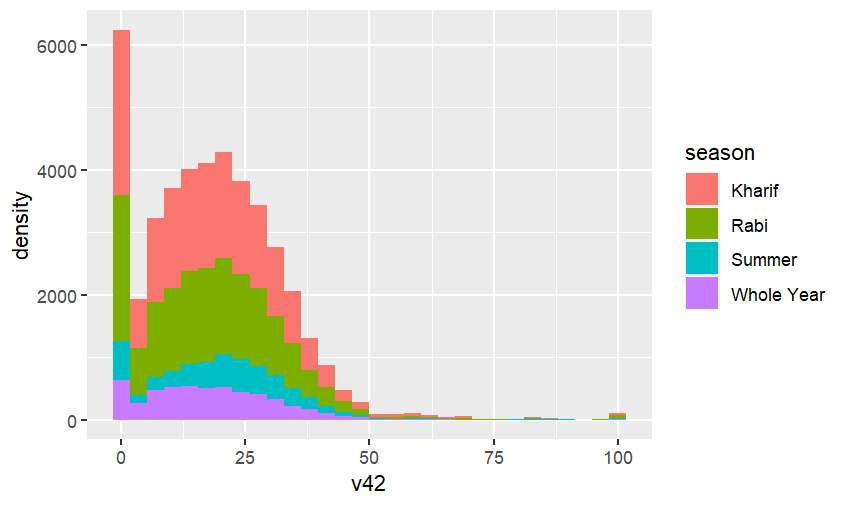
Colored graph -> Data for different seasons with the index on the right



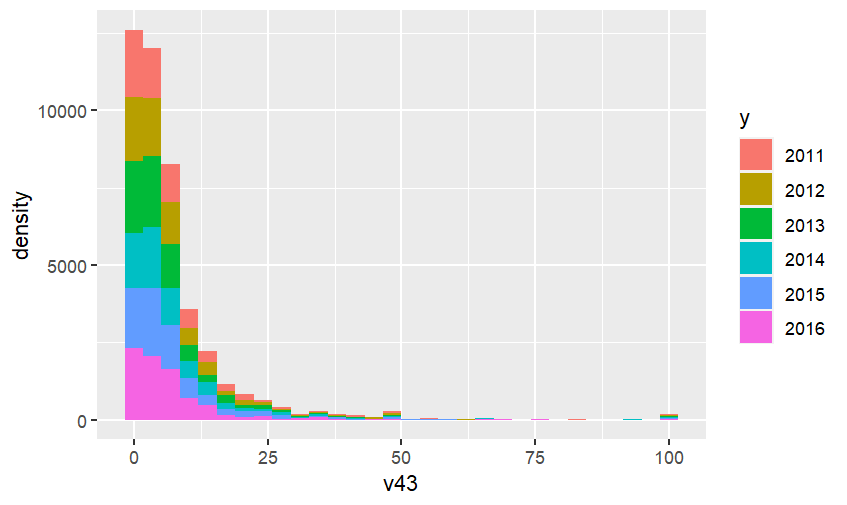
Year-wise Histogram for the percentage of infant deaths from low birth weight.  
x axis -> percentage of infant deaths from low birth weight   
y axis -> frequency of percentage of infant deaths from low birth weight lying in the bin regions



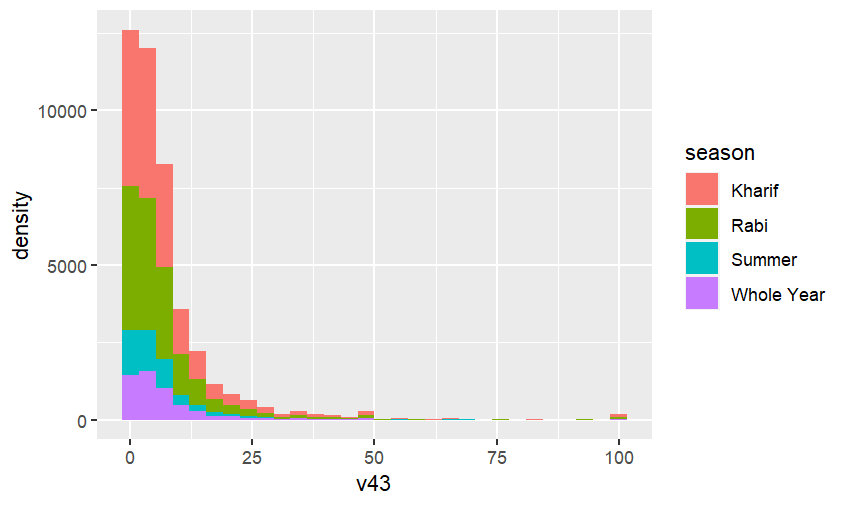
Season-wise Histogram for the percentage of infant deaths from low birth weight.  
x axis -> percentage of infant deaths from low birth weight   
y axis -> frequency of percentage of infant deaths from low birth weight lying in the bin regions



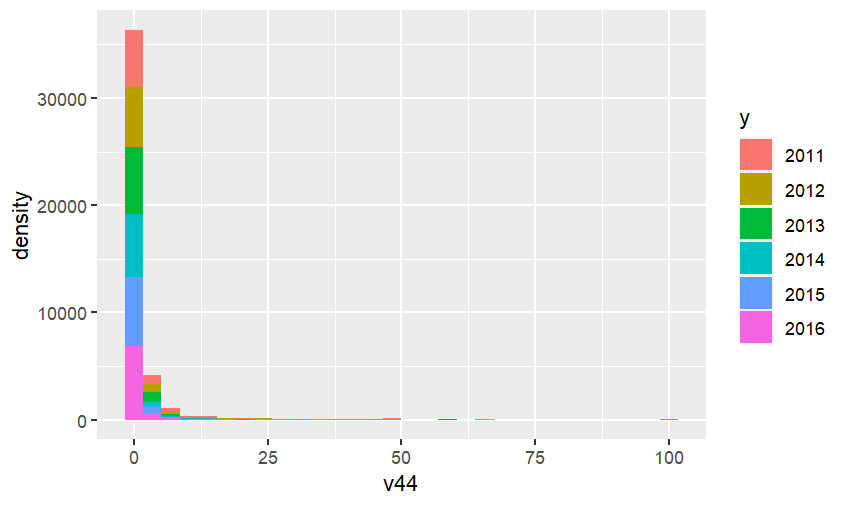
Year-wise Histogram for the percentage of infant deaths from pneumonia.  
x axis -> percentage of infant deaths from pneumonia  
y axis -> frequency of percentage of infant deaths from pneumonia lying in the bin regions



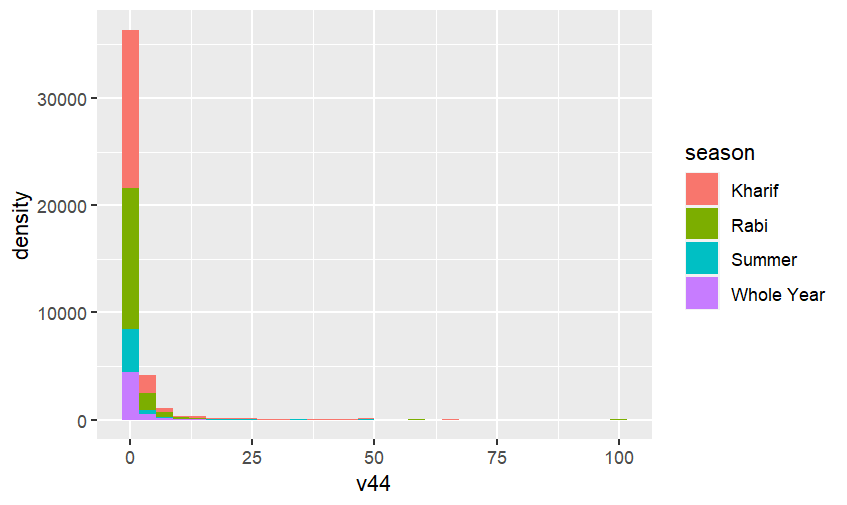
Season-wise Histogram for the percentage of infant deaths from pneumonia.  
x axis -> percentage of infant deaths from pneumonia  
y axis -> frequency of percentage of infant deaths from pneumonia lying in the bin regions



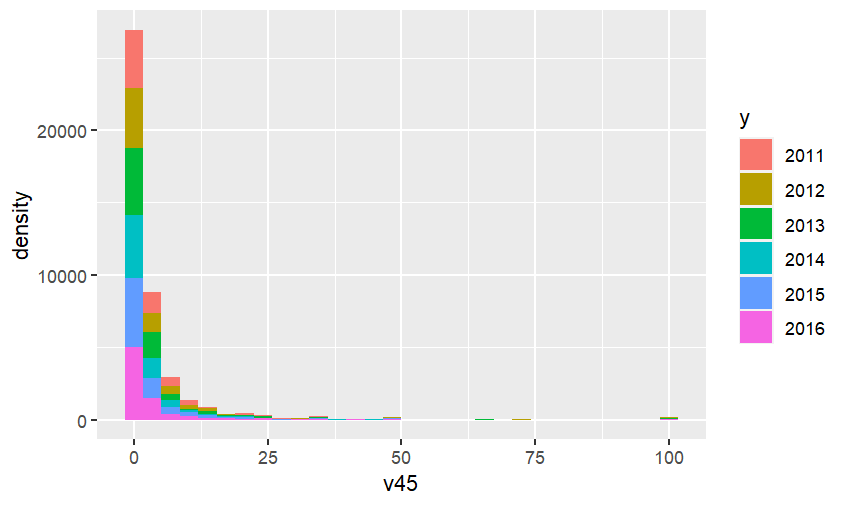
Year-wise Histogram for the percentage of infant deaths from diarrhea.  
x axis -> percentage of infant deaths from diarrhea   
y axis -> frequency of percentage of infant deaths from diarrhea lying in the bin regions



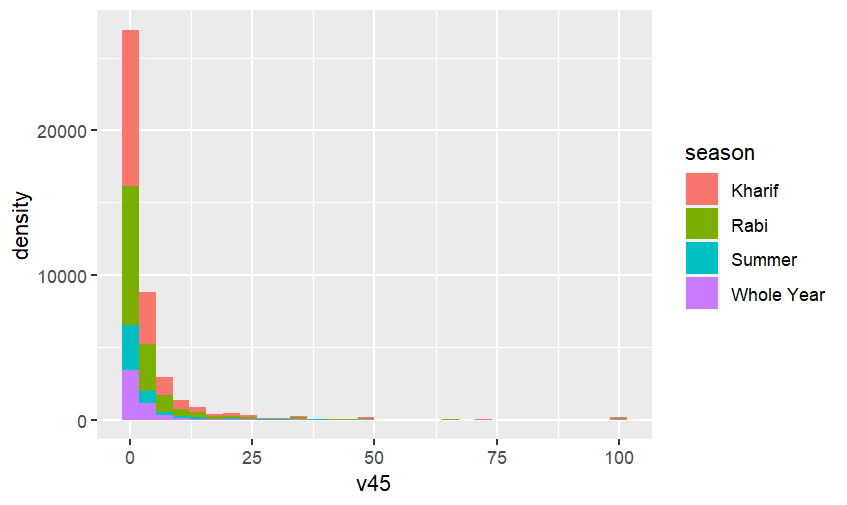
Season-wise Histogram for the percentage of infant deaths from diarrhea.  
x axis -> percentage of infant deaths from diarrhea  
y axis -> frequency of percentage of infant deaths from diarrhea lying in the bin regions



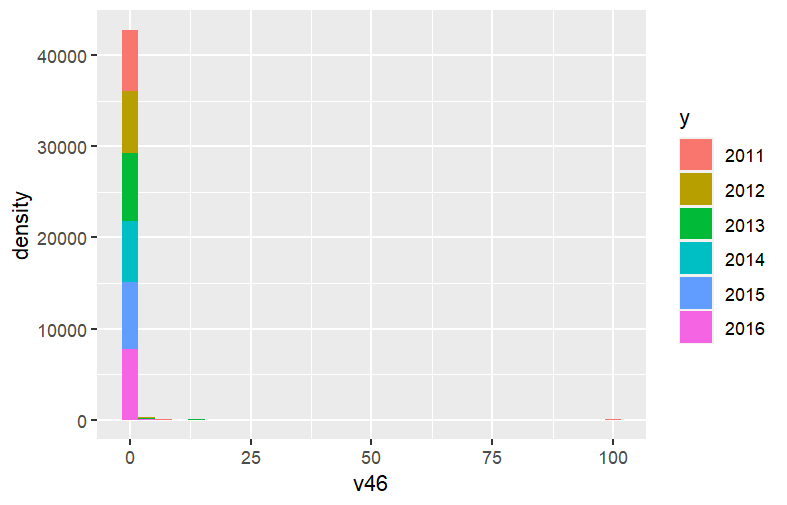
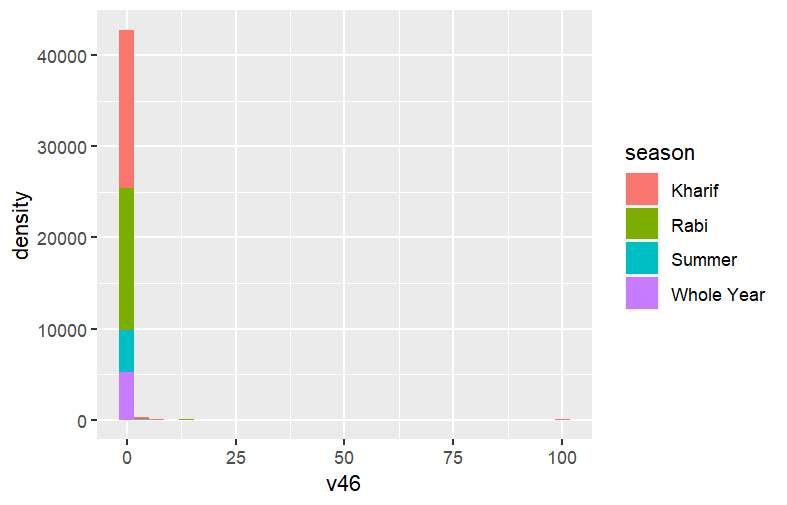
Year-wise Histogram for the percentage of infant deaths from fever.  
x axis -> percentage of infant deaths from fever  
y axis -> frequency of percentage of infant deaths from fever lying in the bin regions



Season-wise Histogram for the percentage of infant deaths from fever.  
x axis -> percentage of infant deaths from fever  
y axis -> frequency of percentage of infant deaths from fever lying in the bin regions



Year-wise Histogram for the percentage of infant deaths from measles.  
x axis -> percentage of infant deaths from measles  
y axis -> frequency of percentage of infant deaths from mesles lying in the bin region

Season-wise Histogram for the percentage of infant deaths from measles.  
x axis -> percentage of infant deaths from measles  
y axis -> frequency of percentage of infant deaths from measles lying in the bin region

C. Outliers in the above graphs are almost same for the yearwise and seasonwise graphs thus  
can be reported for year and season at once for each graph. Given are the values greater than the given base value are outlier

v40 -> 40

v42 -> 50

v43 -> 40

v44 -> 15

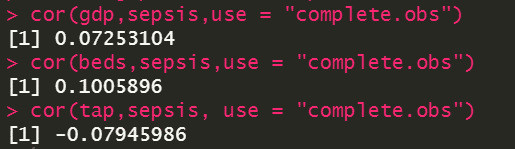
v45 -> 25

v46 -> 5

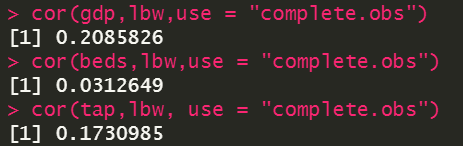
This can be cearly seen if plotted on graph individually for a single year as we can see every year follow the same outlier

D.

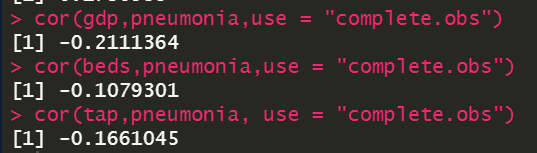
1. Correlation between dependent variable(%age of infants death from sepsis) and the 3 explanatory variables.



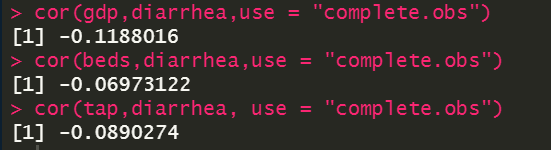
Correlation between dependent variable(%age of infants death from low birth weight) and the 3 explanatory variables.



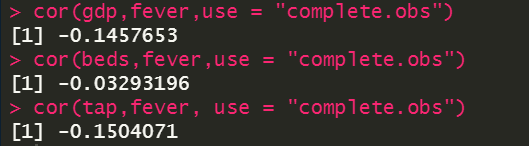
Correlation between dependent variable(%age of infants death from pneumonia) and the 3 explanatory variables.



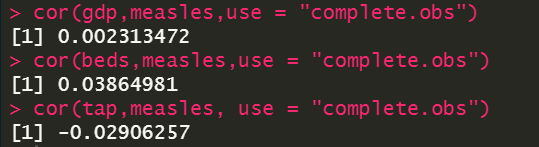
Correlation between dependent variable(%age of infants death from diarrhea) and the 3 explanatory variables.



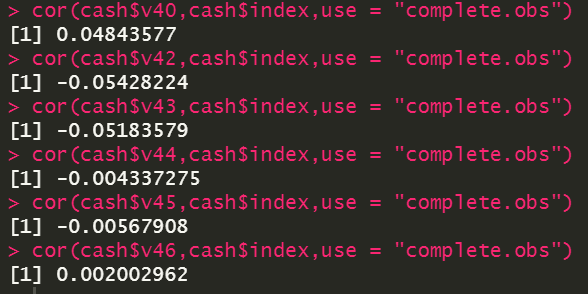
Correlation between dependent variable(%age of infants death from fever) and the 3 explanatory variables.



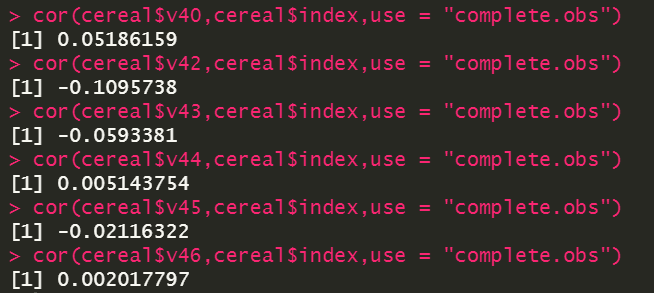
Correlation between dependent variable(%age of infants death from measles) and the 3 explanatory variables.



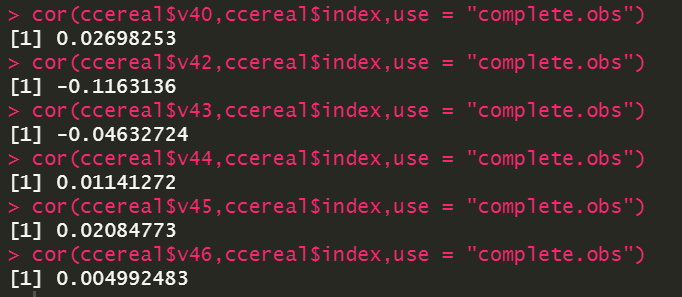
1. Correlation between the yield indices of Cash Crops and 6 DV’s



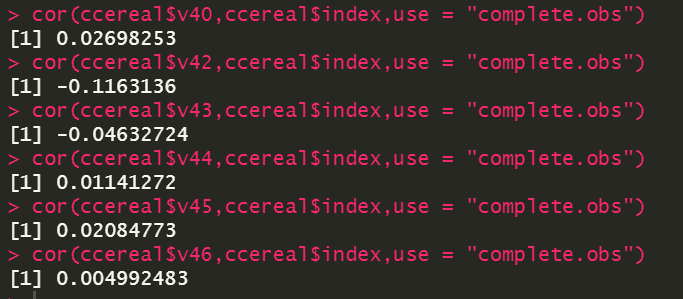
Correlation between the yield indices of Cereal Crops and 6 DV’s



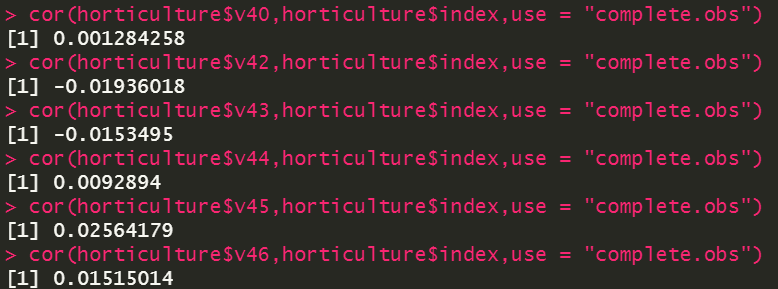
Correlation between the yield indices of Coarse Cereal Crops and 6 DV’s



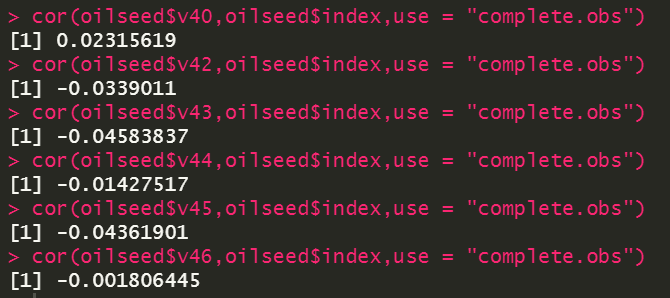
Correlation between the yield indices of Pulse Crop and 6 DV’s



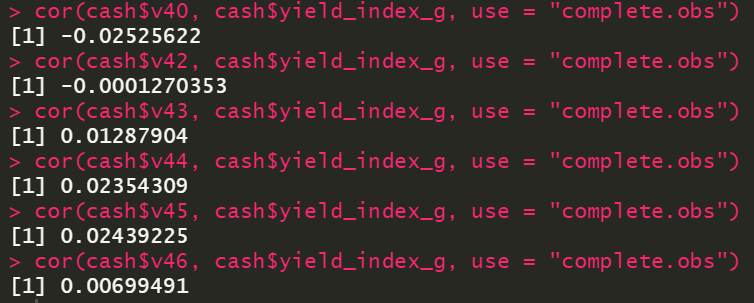
Correlation between the yield indices of Horticulture Crop and 6 DV’s



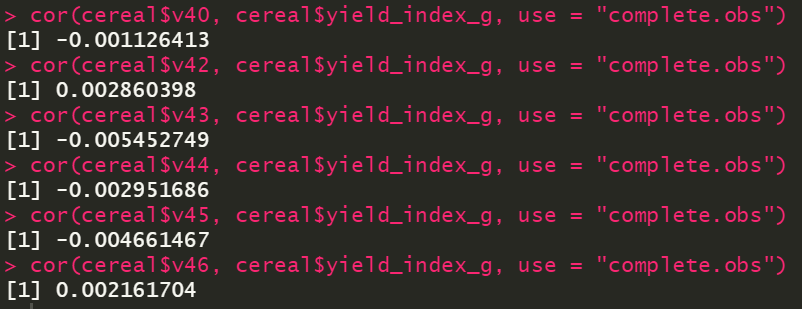
Correlation between the yield indices of Oilseed Crop and 6 DV’s



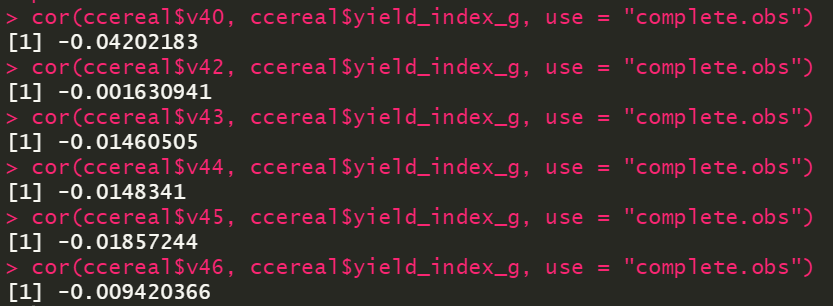
1. Correlation between the yield index growth of Cash Crops and 6 DV’s



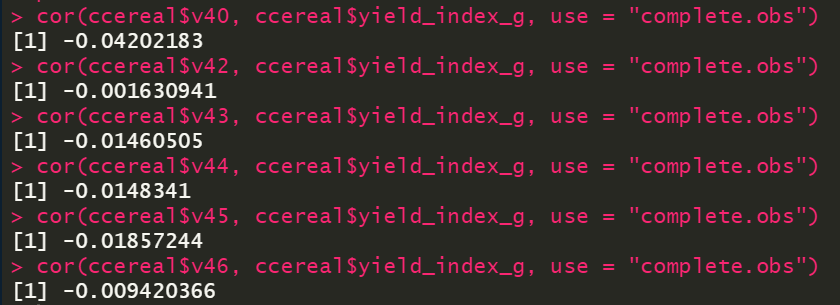
Correlation between the yield index growth of Cereal Crops and 6 DV’s



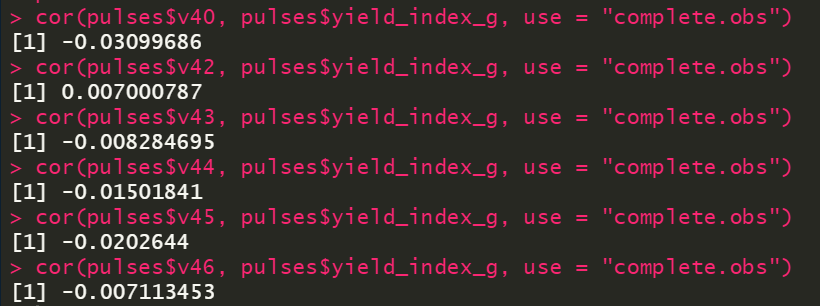
Correlation between the yield index growth of Coarse Cereal Crops and 6 DV’s



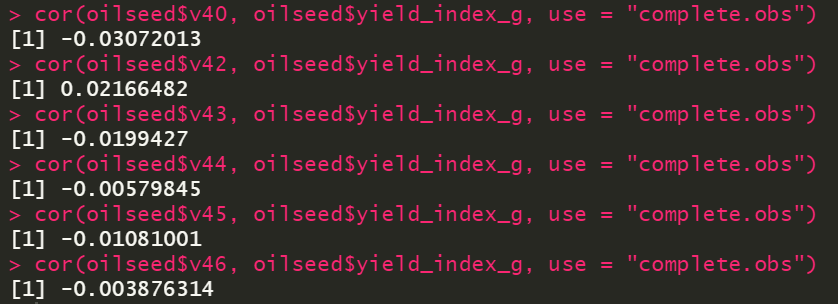
Correlation between the yield index growth of Coarse Cereal Crops and 6 DV’s

****

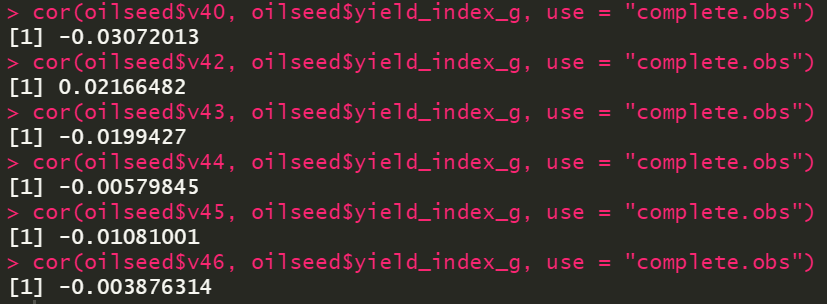
Correlation between the yield index growth of Pulse Crop and 6 DV’s



Correlation between the yield index growth of Horticulture and 6 DV’s



Correlation between the yield index growth of Oilseed crops and 6 DV’s



**Q3.**

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.126e+00  (7.050e-02)\*\*\* |
| **gdp** | 7.082e-07  (1.821e-07) \*\*\* |
| **beds** | 6.760e-06  (7.361e-07) \*\*\* |
| **tap** | -2.924e-02  (1.709e-03) \*\*\* |
| **N = 70568 | Multiple R2 =** 0.01731 **Adjusted R2 =** 0.01724 | |

Crop-wise regression model

Pulses

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.665e+00  (1.933e-01) \*\*\* |
| **gdp** | 2.285e-07  (4.408e-07) |
| **beds** | 6.080e-06  (1.753e-06) \*\*\* |
| **tap** | -3.266e-02  (4.015e-03) \*\*\* |
| **index** | -6.733e-02  (1.101e-01) |
| **N = 13671 | Multiple R2 =** 0.01562 **Adjusted R2 =** 0.01513 | |

Cash

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.083e+00  (1.918e-01) \*\*\* |
| **gdp** | 2.023e-06  (4.704e-07) \*\*\* |
| **beds** | 3.943e-06  (1.959e-06) \* |
| **tap** | -4.092e-02  (4.352e-03) \*\*\* |
| **index** | 4.940e-03  (3.069e-03) |
| **N = 9523 | Multiple R2 =** 0.0293 **Adjusted R2 =** 0.02863 | |

Coarse Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.185e+00  (2.616e-01) \*\*\* |
| **gdp** | 8.167e-07  (4.842e-07). |
| **beds** | 4.479e-06  (2.014e-06) \* |
| **tap** | -4.021e-02  (4.747e-03) \*\*\* |
| **index** | 2.976e-01  (1.282e-01) \* |
| **N = 8890 | Multiple R2 =** 0.02279 **Adjusted R2 =** 0.02205 | |

Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 4.459e+00  (2.300e-01) \*\*\* |
| **gdp** | 2.300e-01  (4.295e-07) |
| **beds** | 6.454e-06  (1.773e-06) \*\*\* |
| **tap** | -2.505e-02  (4.180e-03) \*\*\* |
| **index** | 3.738e-01  (7.616e-02) \*\*\* |
| **N = 14246 | Multiple R2 =** 0.01303 **Adjusted R2 =** 0.01255 | |

Horticulture

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 4.931e+00  (1.838e-01) \*\*\* |
| **gdp** | 7.304e-07  (4.767e-07) |
| **beds** | 8.199e-06  (1.727e-06) \*\*\* |
| **tap** | -2.187e-02  (4.245e-03) \*\*\* |
| **index** | -1.867e-02  (1.084e-02). |
| **N = 12369 | Multiple R2 =** 0.018 **Adjusted R2 =** 0.01746 | |

Oilseeds

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.103e+00  (1.840e-01) \*\*\* |
| **gdp** | 1.670e-07  (4.159e-07) |
| **beds** | 7.373e-06  (1.764e-06) \*\*\* |
| **tap** | -3.073e-02  (4.074e-03) \*\*\* |
| **index** | 2.405e-01  (7.635e-02) \*\* |
| **N = 11605 | Multiple R2 =** 0.01834 **Adjusted R2 =** 0.01777 | |

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.167e+00  (8.138e-02) \*\*\* |
| **gdp** | 5.664e-07  (1.839e-07) \*\* |
| **beds** | 6.471e-06  (7.417e-07) \*\*\* |
| **tap** | -3.052e-02  (1.735e-03) \*\*\* |
| **index cereal** | 1.174e-01  (3.868e-02) \*\* |
| **index coarse cereal** | 1.295e-01  (7.705e-02) . |
| **index cash** | 8.818e-03  (2.714e-03) \*\* |
| **index oilseed** | 1.605e-01  (6.296e-02) \* |
| **index horticulture** | -4.424e-03  (9.258e-03) |
| **index pulses** | 7.416e-02  (8.599e-02) |
| **N = 70562 | Multiple R2 =** 0.01709 **Adjusted R2 =** 0.01688 | |

1. Pulses

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.657e+00  (1.899e-01) \*\*\* |
| **gdp** | 5.866e-07  (4.830e-07) |
| **beds** | 4.341e-06  (1.932e-06) \* |
| **tap** | -3.251e-02  (4.476e-03) \*\*\* |
| **Index growth** | -1.506e-01  (7.522e-02) \* |
| **N = 13671 | Multiple R2 =** 0.01504 **Adjusted R2 = 0.01444** | |

Cash

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.111e+00  (2.095e-01) \*\*\* |
| **gdp** | 2.403e-06  (5.284e-07)\*\*\* |
| **beds** | 3.626e-06  (2.224e-06) |
| **tap** | -3.998e-02  (5.049e-03)\*\*\* |
| **Index growth** | -4.337e-02  (3.683e-02) |
| **N = 9523 | Multiple R2 =** 0.03054 **Adjusted R2 = 0.0297** | |

Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.188e+00  (1.795e-01) \*\*\* |
| **gdp** | 2.951e-07  (4.636e-07)\*\*\* |
| **beds** | 6.379e-06  (1.929e-06) |
| **tap** | -1.907e-02  (4.571e-03)\*\*\* |
| **Index growth** | -5.071e-06  (7.684e-05) |
| **N = 14264 | Multiple R2 =** 0.009036 **Adjusted R2 = 0.008468** | |

Coarse Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.511e+00  (2.468e-01) \*\*\* |
| **gdp** | 1.006e-06  (5.321e-07). |
| **beds** | 3.925e-06  (2.199e-06). |
| **tap** | -3.673e-02  (5.293e-03)\*\*\* |
| **Index growth** | -1.853e-01  (7.959e-02)\* |
| **N = 8890 | Multiple R2 =** 0.02145 **Adjusted R2 = 0.02054** | |

Horticulture

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 4.907e+00  (2.025e-01) \*\*\* |
| **gdp** | 1.334e-06  (5.559e-07)\* |
| **beds** | 5.762e-06  (1.966e-06)\*\* |
| **tap** | -2.344e-02  (4.898e-03)\*\*\* |
| **Index growth** | -5.594e-03  (8.519e-03) |
| **N = 12589 | Multiple R2 =** 0.01678 **Adjusted R2 = 0.01611** | |

Oilseeds

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.298e+00  (1.860e-01) \*\*\* |
| **gdp** | 5.996e-07  (4.490e-07) |
| **beds** | 5.575e-06  (1.931e-06)\*\* |
| **tap** | -2.677e-02  (4.479e-03)\*\*\* |
| **Index growth** | -7.100e-02  (3.432e-02)\* |
| **N = 11605 | Multiple R2 =** 0.01612 **Adjusted R2 = 0.01542** | |

E.

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.167e+00  (8.138e-02) \*\*\* |
| **gdp** | 5.664e-07  (1.839e-07)\*\* |
| **beds** | 6.471e-06  (7.417e-07)\*\*\* |
| **tap** | -3.052e-02  (1.735e-03)\*\* |
| **index growth cereal** | 1.174e-01  (3.868e-02). |
| **index growth coarse cereal** | 1.295e-01  (7.705e-02)\*\* |
| **index growth cash** | 8.818e-03  (2.714e-03)\* |
| **index growth oilseed** | 1.605e-01  (6.296e-02) |
| **index growth horticulture** | -4.424e-03  (9.258e-03) |
| **index growth pulses** | 7.416e-02  (8.599e-02) |
| **N = 70562 | Multiple R2 =** 0.01709 **Adjusted R2 = 0.01688** | |

F.

Pulses

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 9.57993  (1.34695) \*\*\* |
| **log(gdp)** | -1.98235  (0.24769)\*\*\* |
| **log(beds)** | 1.99287  (0.21395)\*\*\* |
| **log(tap)** | -0.17652  (0.07013)\* |
| **log(Index growth)** | -0.10115  (0.34744) |
| **N = 13671 | Multiple R2 =** 0.01306 **Adjusted R2 =** 0.01256 | |

Cash

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 3.46551  (1.41918) \* |
| **log(gdp)** | -1.50208  (0.26950)\*\*\* |
| **log(beds)** | 2.01118  (0.23746)\*\*\* |
| **log(tap)** | -0.33250  (0.07582)\*\*\* |
| **log(Index growth)** | 0.07428  (0.06663) |
| **N = 9523 | Multiple R2 =** 0.02037 **Adjusted R2 = 0.0197** | |

Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 7.53917  (1.22805) \*\*\* |
| **log(gdp)** | -1.87803  (0.24376)\*\*\* |
| **log(beds)** | 1.91193  (0.21673)\*\*\* |
| **log(tap)** | -0.10281  (0.06956) |
| **log(Index growth)** | 1.10026  (0.25995)\*\*\* |
| **N = 14264 | Multiple R2 =** 0.01279 **Adjusted R2 =** 0.01232 | |

Coarse Cereal

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 5.30262  (1.70657) \*\* |
| **log(gdp)** | -1.31387  (0.30329)\*\*\* |
| **log(beds)** | 1.58772  (0.25652)\*\*\* |
| **log(tap)** | -0.37654  (0.08878)\*\*\* |
| **log(Index growth)** | 0.08878  (0.32092)\* |
| **N = 8890 | Multiple R2 =** 0.01532 **Adjusted R2 =** 0.01458 | |

Horticulture

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 7.15724  (1.27821) \*\*\* |
| **gdp** | -2.08757  (0.24319)\*\*\* |
| **beds** | 2.29034  (0.21067)\*\*\* |
| **tap** | 0.04813  (0.07232) |
| **Index growth** | -0.15612  (0.09793) |
| **N = 12589 | Multiple R2 =** 0.01799 **Adjusted R2 = 0.01746** | |

Oilseeds

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 8.04257  (1.21201) \*\*\* |
| **log(gdp)** | -2.09674  (0.24518)\*\*\* |
| **log(beds)** | 2.15241  (0.21778)\*\* |
| **log(tap)** | -0.23492  (0.07185)\*\*\* |
| **log(Index growth)** | 1.77341  (0.29121)\*\*\* |
| **N = 11605 | Multiple R2 =** 0.02104 **Adjusted R2 = 0.02046** | |

G.

|  |  |
| --- | --- |
| **sepsis (v40)** | **Model 1** |
| **Coefficient (SE)** |
| **Intercept** | 7.13292  (0.54720) \*\*\* |
| **gdp** | -1.83732  (0.10460) \*\*\* |
| **beds** | 2.00525  (0.09138) \*\*\* |
| **tap** | -0.16121  (0.03017) \*\* |
| **log(index cereal)** | 0.30499  (0.10174). \* |
| **log(Index coarse cereal)** | 0.37939  (0.16250) \*\*\* |
| **log(index cash)** | 0.17443  (0.04755) \*\*\* |
| **log(index oilseed)** | 0.61631  (0.16627) \*\*\* |
| **log(index horticulture)** | -0.05603  (0.06849) |
| **log(index pulses)** | 0.53285  (0.20155) \*\* |
| **N = 70562 | Multiple R2 =** 0.01457**Adjusted R2 = 0.01436** | |

**Q4.** We get the R squared value after running regression model on the dataset. The sum of the square of correlation coefficient equals goodness of fit.

**Q5.** As we know that not every crop is grown in every district thus we must take our dadatset as subset of the crops and then run the regression and not take them together.

**Q6.** We can comment on the relation by looking at the correlation between yield index growth and Health indicator for different crop categories as done in q2 d 3.

V40-> Similar across different crop categories at least we can that it is negative indicating the decrease in sepis deaths with increase in yield growth of any crop.

This is the least similarity that can be found thus taking consideration of all the health indicators we can say that there is no clear relationship between the heath inidicators and yield growth across crop categories.