Eco Assignemnt_1_DHEERAJ_2020194

Q1) csv is attached

Q2 part B

V40

mean	5.566724
median	3.1
mode	10
sd	8.297

V42

mean	18.43
median	17.5
mode	13.3
sd	13.9

V43

mean	7.148
median	4.2
mode	30
sd	10.850

V44

mean	1.686
median	0
mode	3.3
sd	6.62

V45

mean	3.822
median	0.9
mode	3.3
sd	9.73077

V 46

mean	0.2
median	0
mode	0
sd	3.134

Q2 part B

#importing ggplot to make histrogram

library(ggplot2)

v40 = dheeraj[,c('v40')]

v42 = dheeraj[,c('v42')]

v43 = dheeraj[,c('v43')]

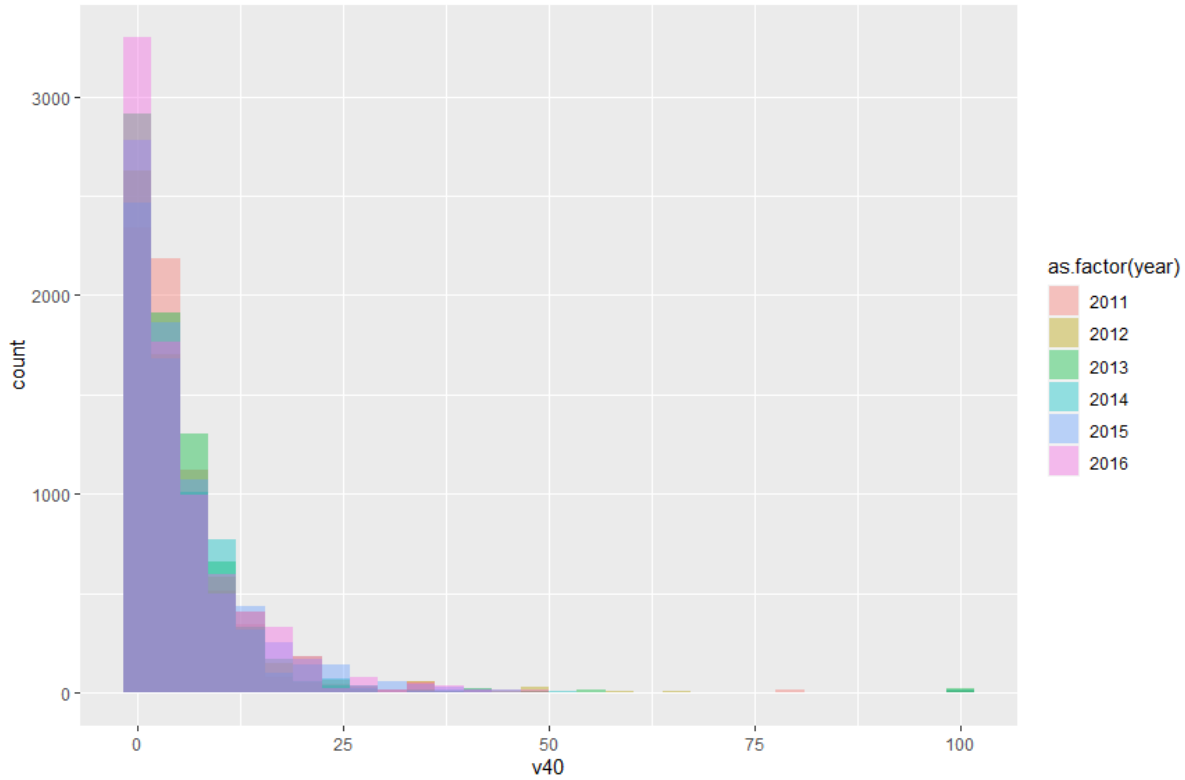
v44 = dheeraj[,c('v44')]

v45 = dheeraj[,c('v45')]

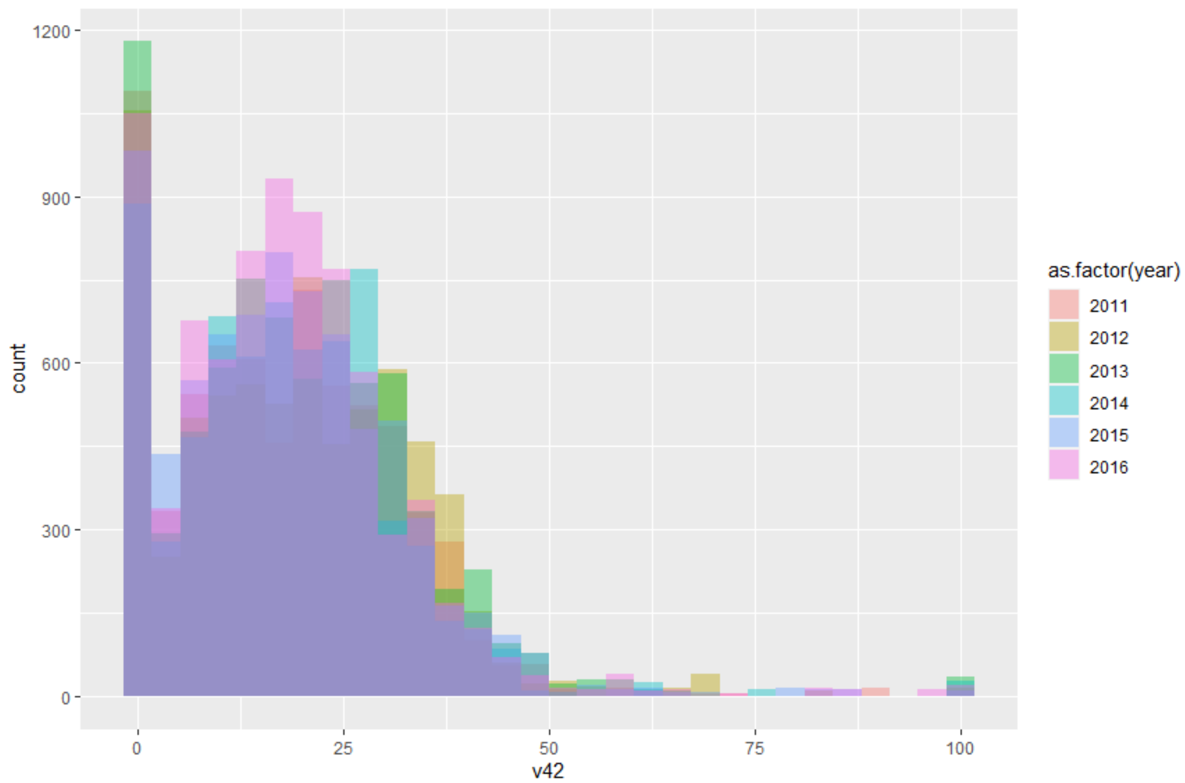
v46 = dheeraj[,c('v46')]

#Year wise

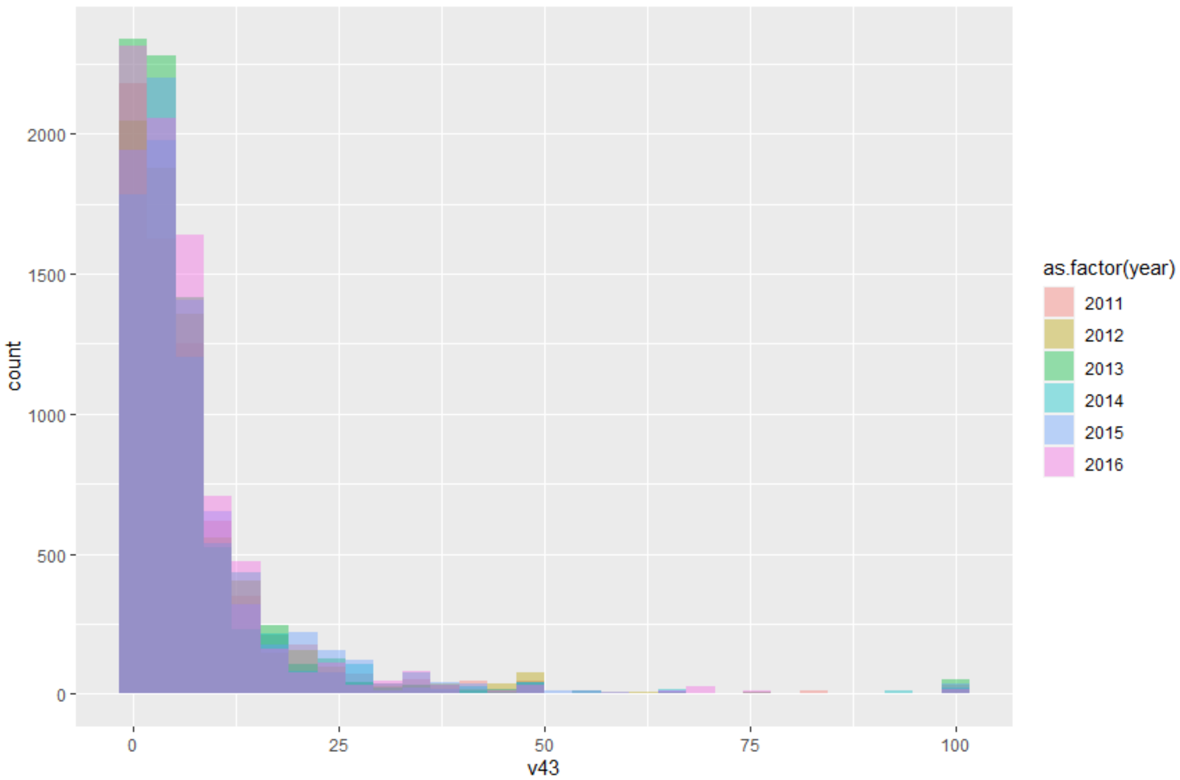
ggplot(dheeraj,aes(v40,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")



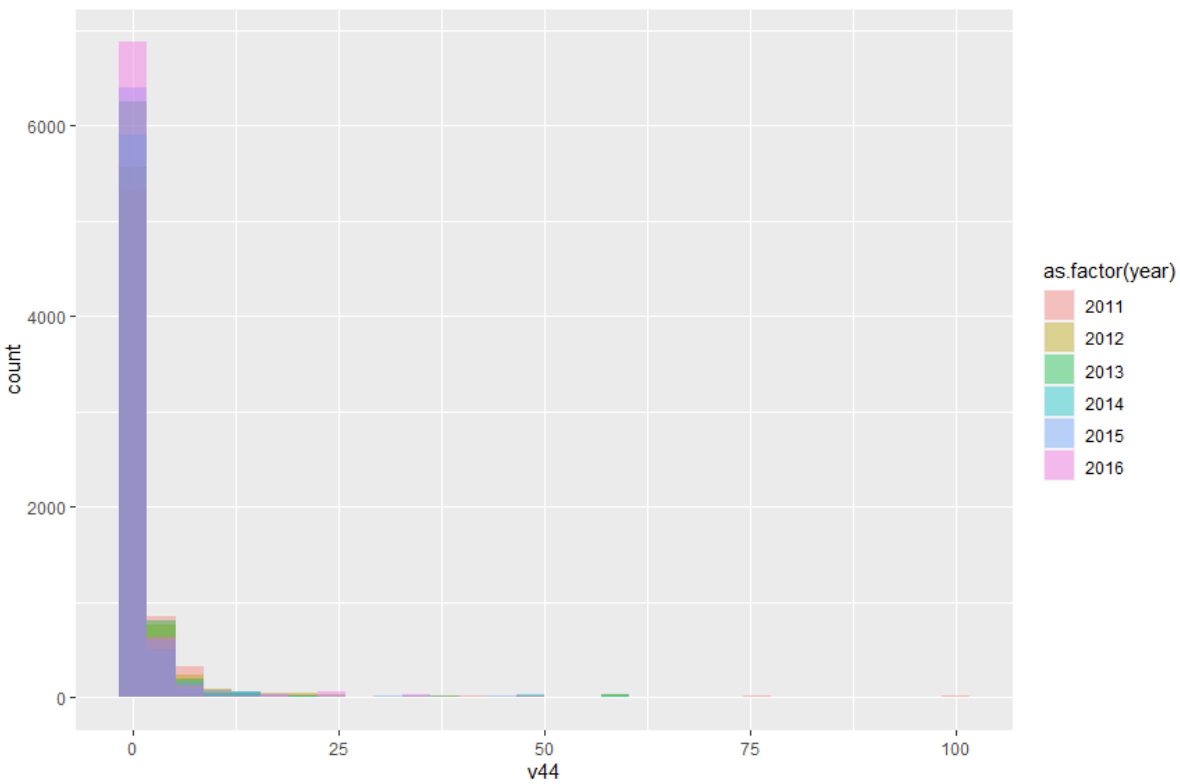
ggplot(dheeraj,aes(v42,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")



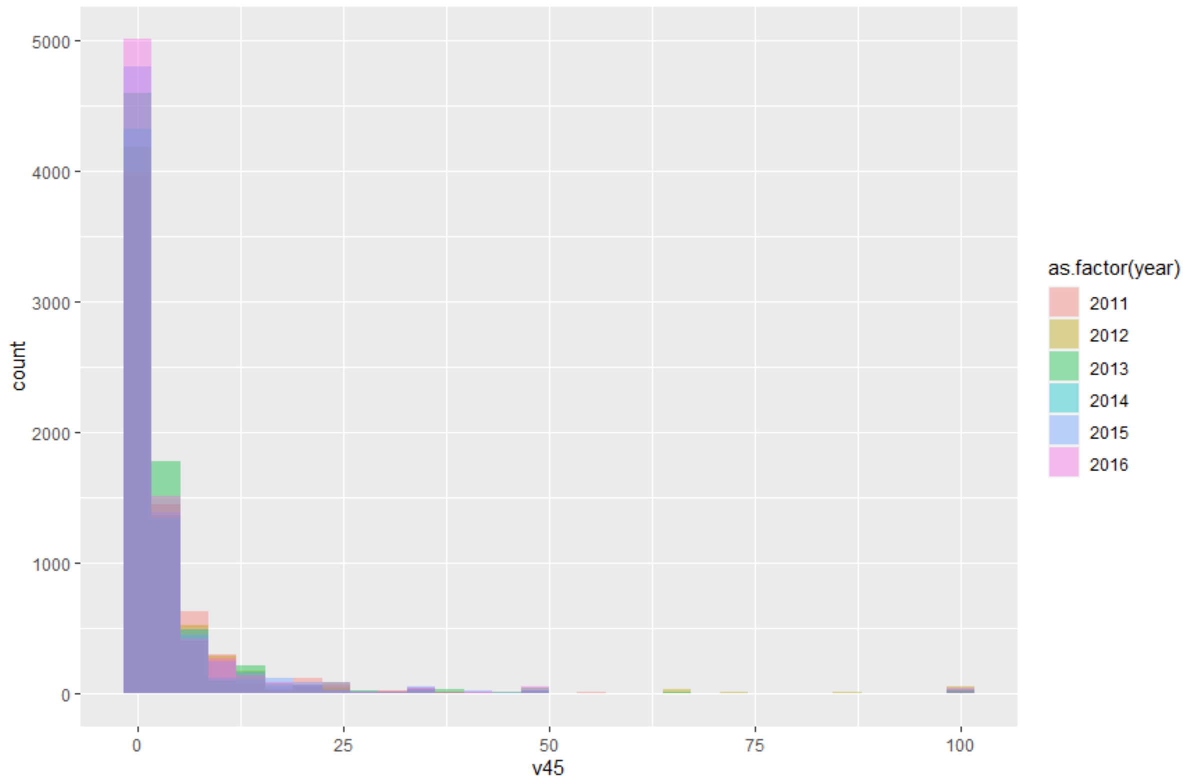
ggplot(dheeraj,aes(v43,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")



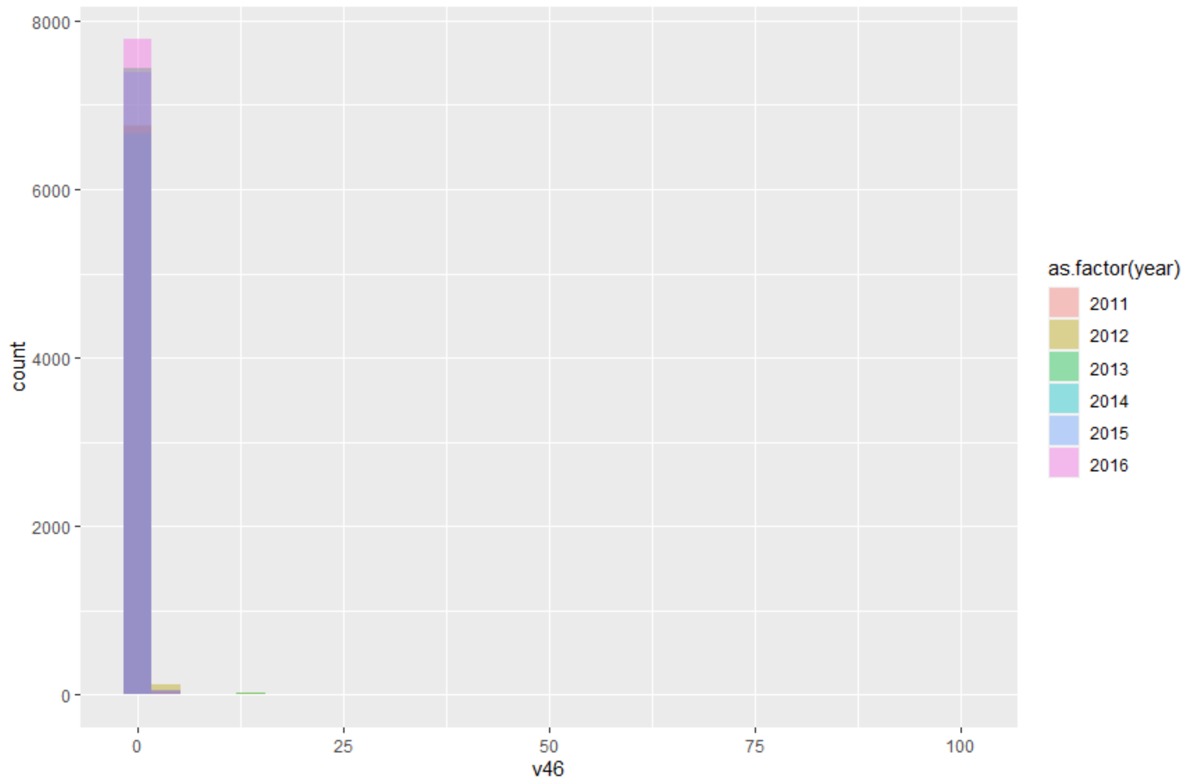

```
ggplot(dheeraj,aes(v44,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")
```




```
ggplot(dheeraj,aes(v45,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")
```

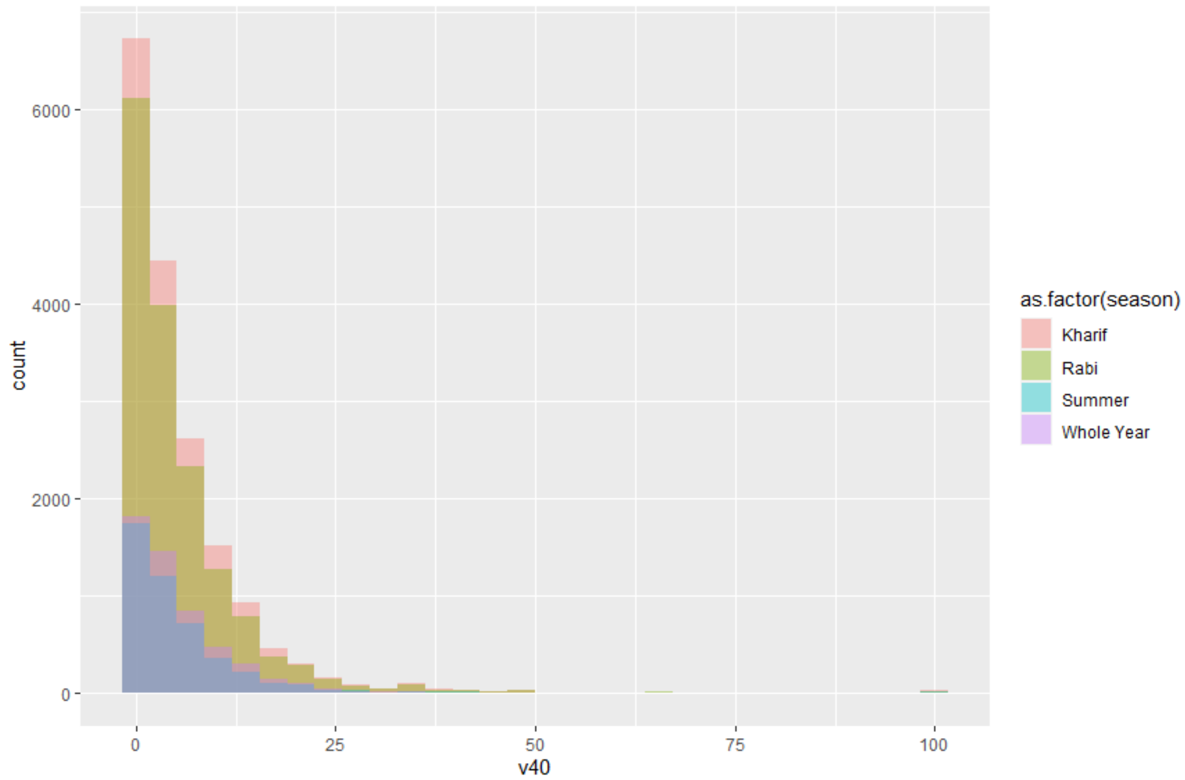


ggplot(dheeraj,aes(v46,fill=as.factor(year)))+geom_histogram(alpha = 0.4,position="identity")

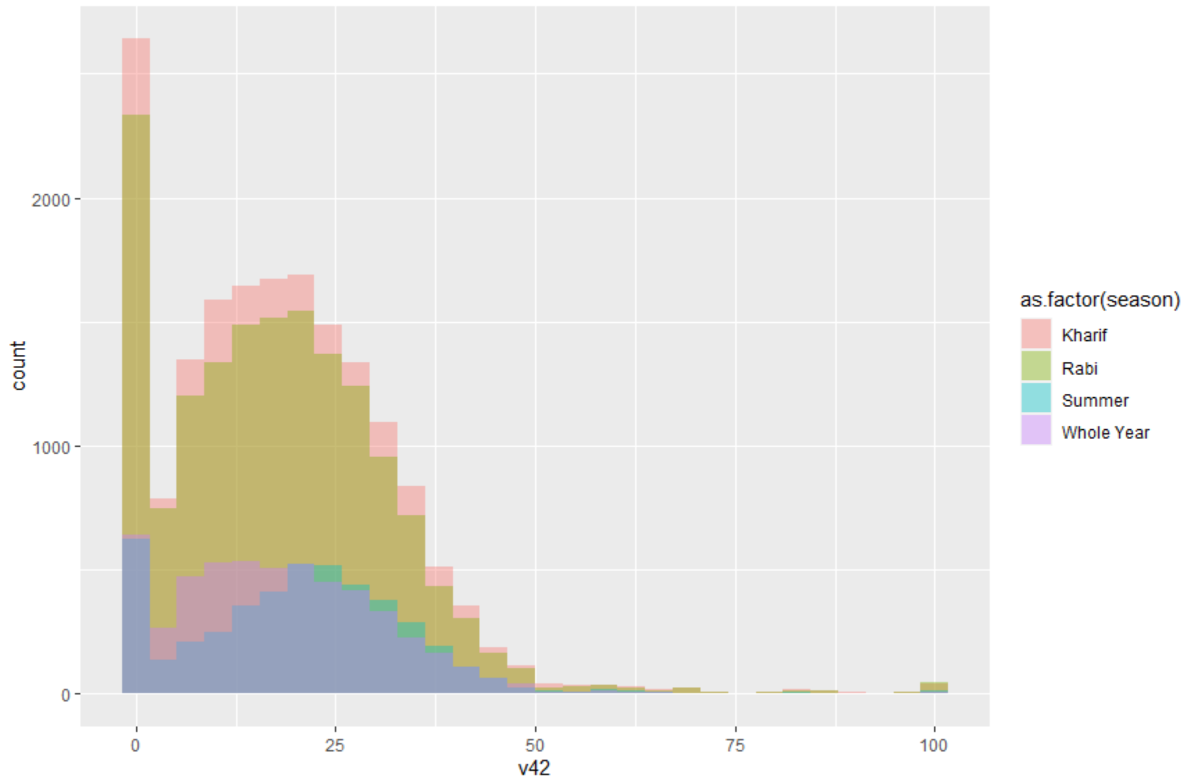


#Season wise

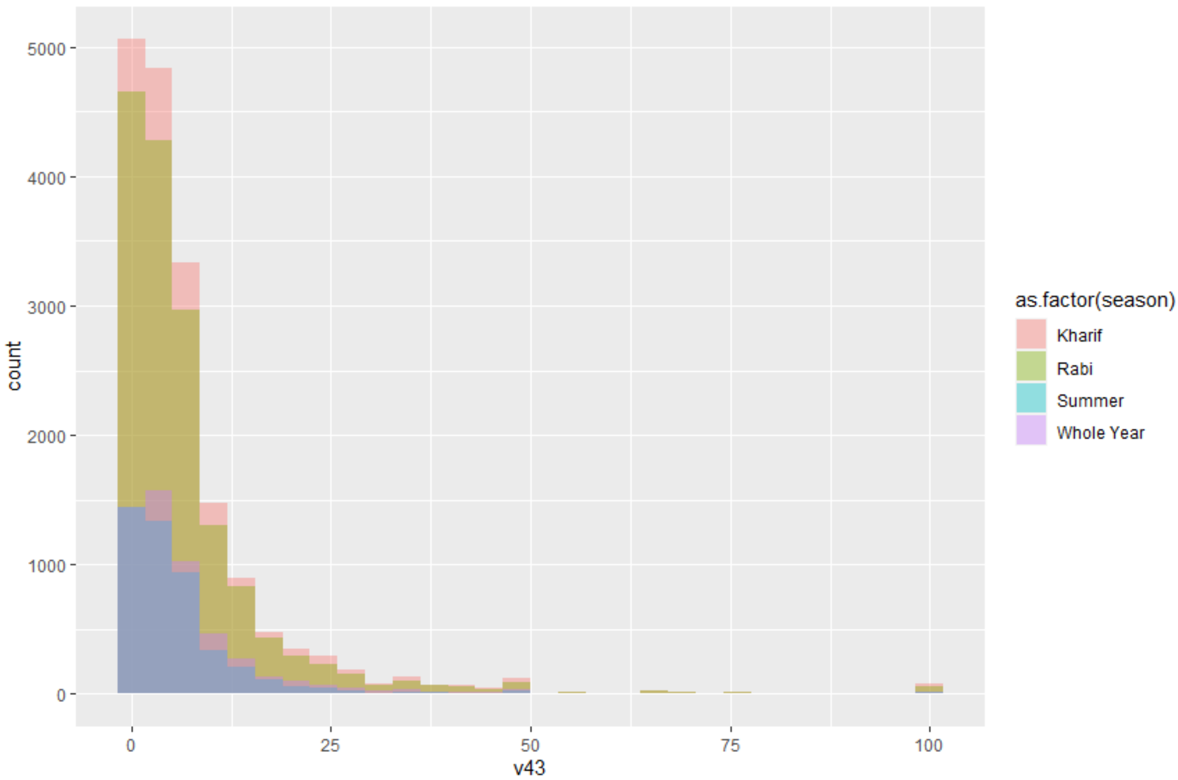
ggplot(dheeraj,aes(v40,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



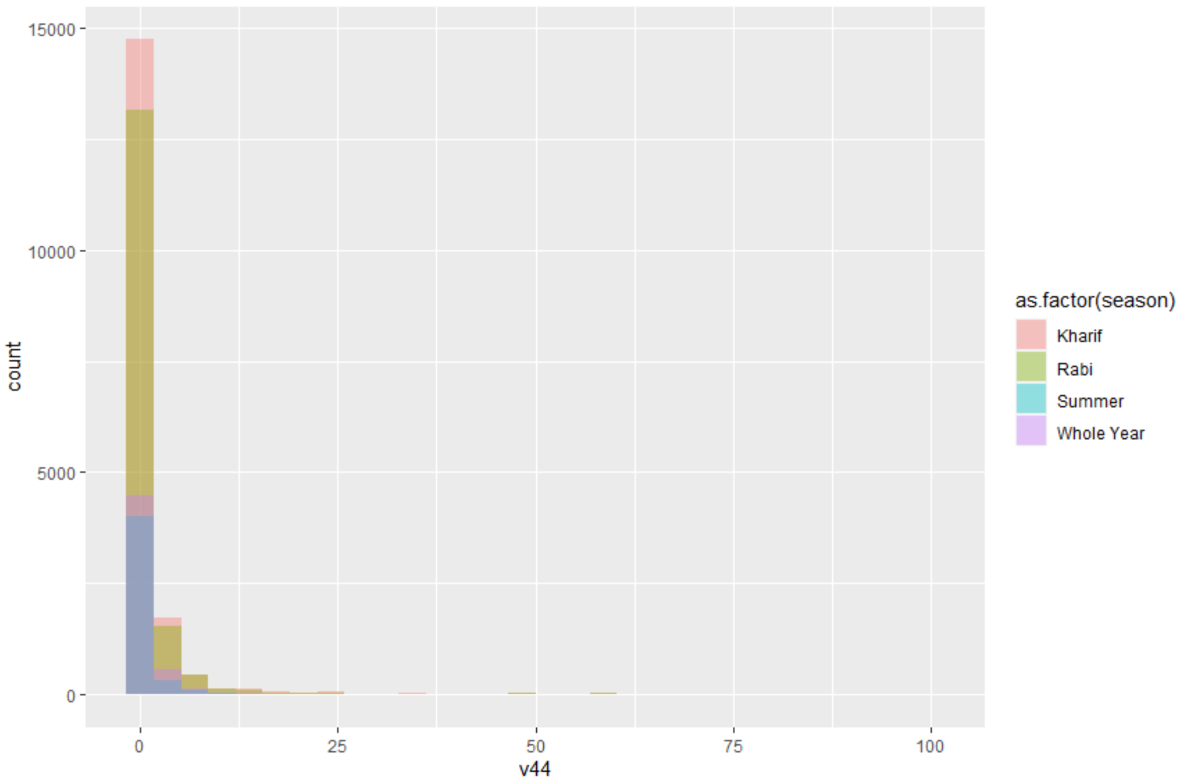
ggplot(dheeraj,aes(v42,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



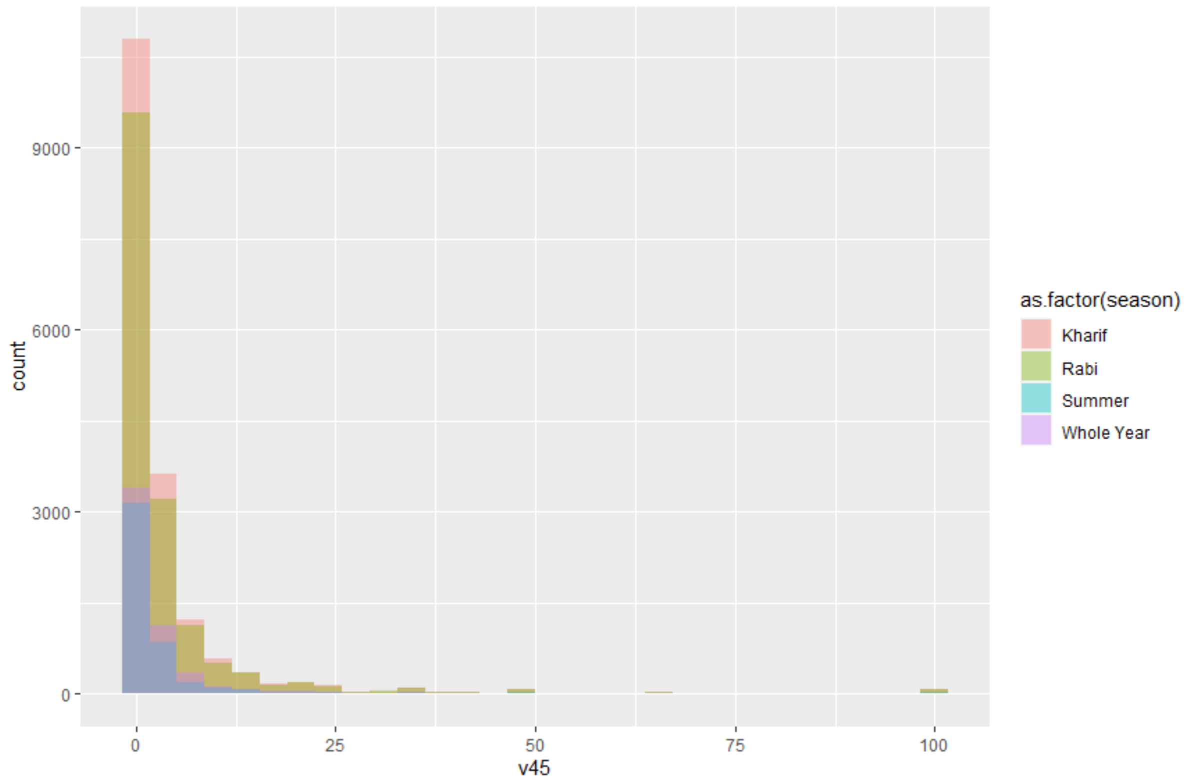
ggplot(dheeraj,aes(v43,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



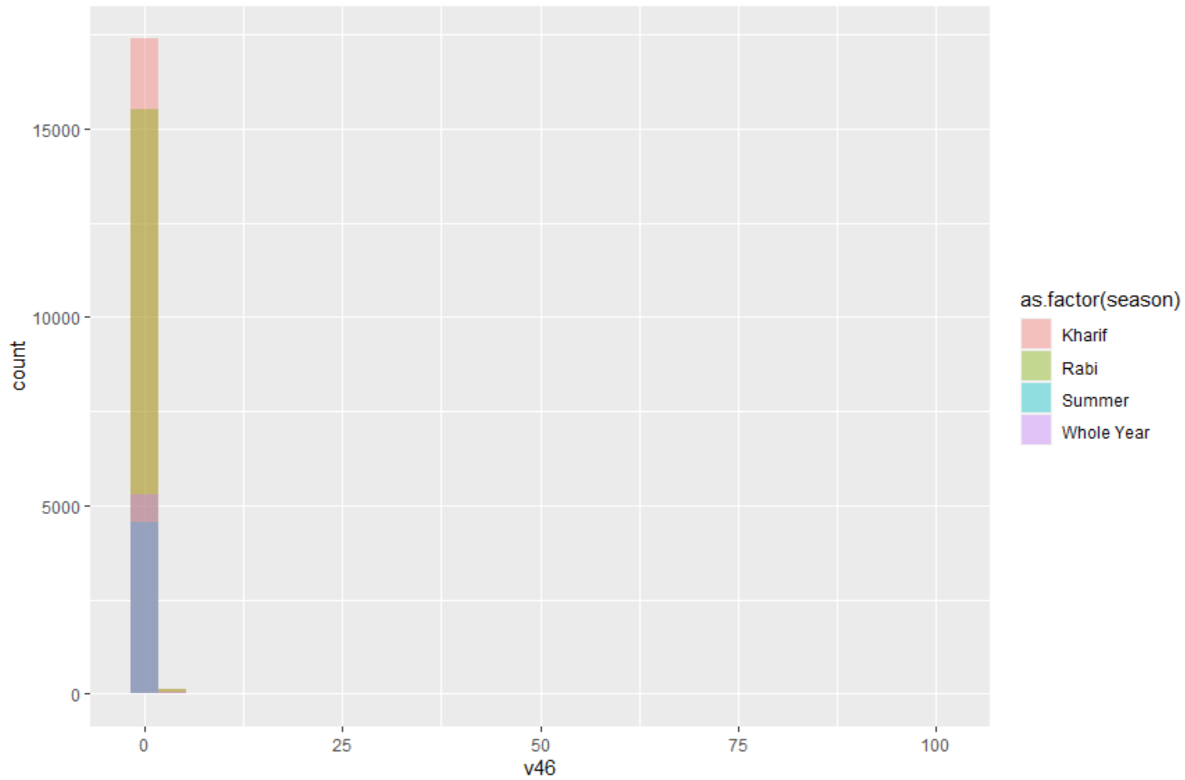
ggplot(dheeraj,aes(v44,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



ggplot(dheeraj,aes(v45,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



ggplot(dheeraj,aes(v46,fill=as.factor(season)))+geom_histogram(alpha = 0.4,position="identity")



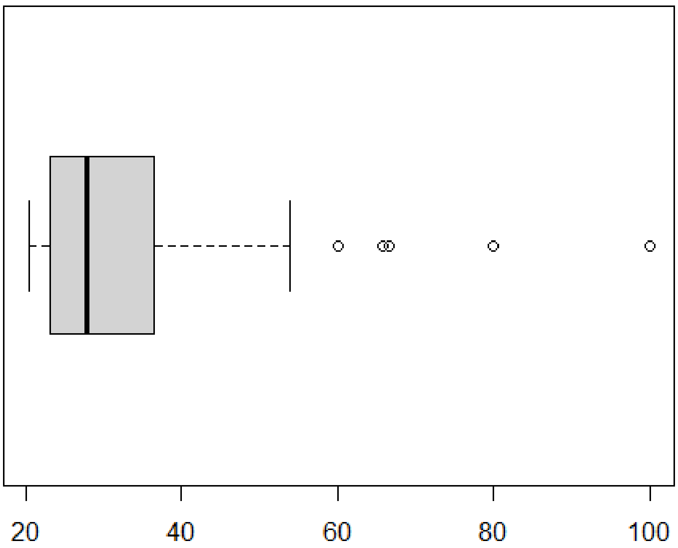
Q2 part C) outliers

The boxplot of each of the following variables are used to compute outliers.

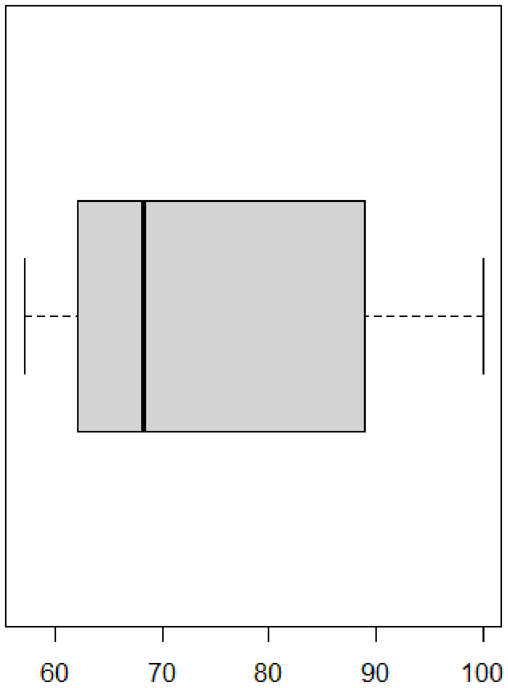
I've listed the range in which the outliers may be found below, and then I've saved them in separate variables.

You can see that in my R scripe

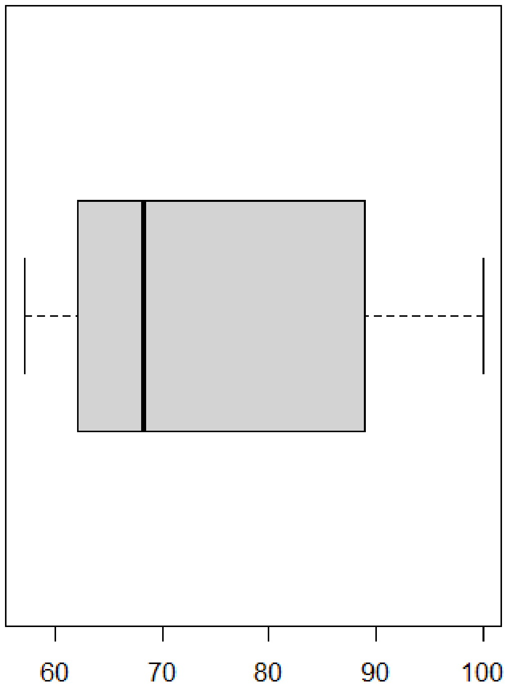
V40



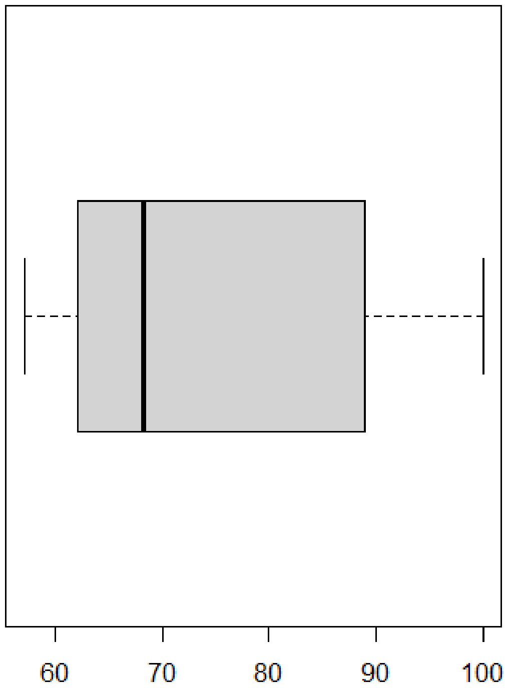
V42



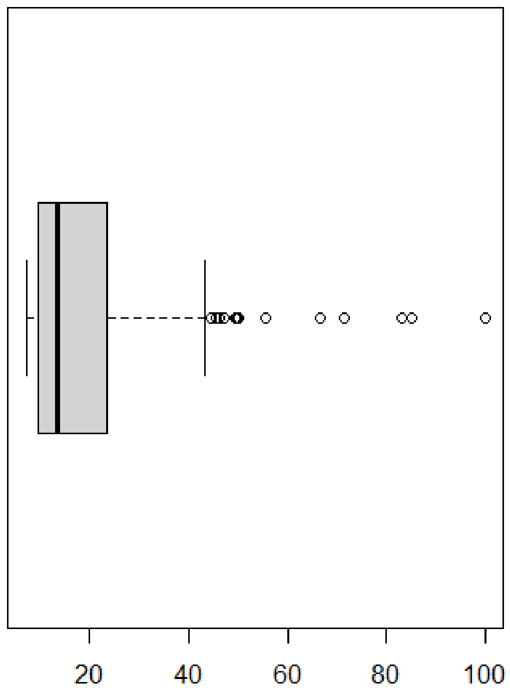
V43



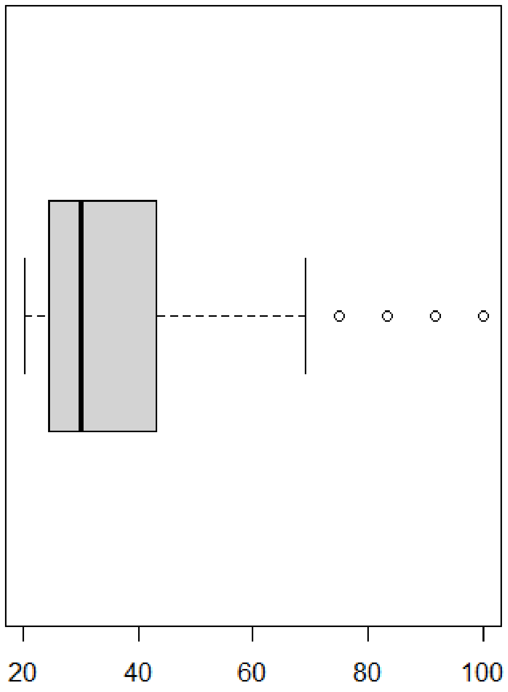
V44



V45



V46



Q2) d) part 1)

V40

gdp	0.0725
tap	-0.080
beds	0.1017

V42

gdp	0.208
tap	0.1725
beds	0.0325

V43

gdp	-0.2111

tap	-0.166
beds	-0.108

v44

gdp	-0.1188
tap	-0.0883
beds	-0.0698

v45

gdp	-0.1457
tap	-0.1502
bedS	-0.0330

v46

gdp	0.00231
tap	-0.029
bedS	0.0386

PART 2)

```
> #2)_d)b)
> #correlation of v40 with cash
> a<-dheeraj$index [dheeraj$cropcategory=='Cash'] # saving cash index
> b<-dheeraj$v40 [dheeraj$cropcategory=='Cash'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] 0.04843577
>
> a<-dheeraj$index [dheeraj$cropcategory=='Cereal'] # saving cash index
> b<-dheeraj$v42 [dheeraj$cropcategory=='Cereal'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] -0.1095738
>
> a<-dheeraj$index [dheeraj$cropcategory=='Horticulture'] # saving cash index
> b<-dheeraj$v43 [dheeraj$cropcategory=='Horticulture'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] -0.0153495
>
> a<-dheeraj$index [dheeraj$cropcategory=='Pulse'] # saving cash index
> b<-dheeraj$v44 [dheeraj$cropcategory=='Pulse'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] 0.02784892
>
> a<-dheeraj$index [dheeraj$cropcategory=='Oilseed'] # saving cash index
> b<-dheeraj$v45 [dheeraj$cropcategory=='Oilseed'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] -0.04361901
>
> a<-dheeraj$index [dheeraj$cropcategory=='Coarse Cereal'] # saving cash index
> b<-dheeraj$v46 [dheeraj$cropcategory=='Coarse Cereal'] # v40 value of cash
> cor(a,b,use = "complete.obs")
[1] 0.004992483
>
```

Q3)

A)

v40	**Model 1**
	Coefficient (SE)
Intercept	5.129
gdp	7.240*10^-09
beds	6.70910^-06
taps	-2.95210^-02

N=73111 | R squared = 0.01743

B)

Pulses

v40	Model 2.1
	Coefficient (SE)
Intercept	5.668
gdp	2.485*10^-09
beds	6.010*10^-06
taps	-3.299*10^-02
index	-6.524*10^-02
N=13676 | R squared = 0.01577	

coarse cereals

v40	Model 2.3
	Coefficient (SE)
Intercept	5.188e
gdp	8.383*10^09

beds	4.405*10^-06
taps	-4.057*10^-02
index	2.993*10^-01
N= 8895 | R squared = 0.023	

cereals

v40	**Model 2.2S**
	Coefficient (SE)
Intercept	4.459
gdp	2.277*10^-09
beds	6.389*10^-06
taps	-2.546*10^-02
index	3.766*10^-01
N= 14269 | R squared = 0.01316	

cash

v40	**Model 2.4**
	Coefficient (SE)

Intercept	5.087
gdp	2.040*10^-08
beds	3.878*10^-06
taps	-4.127*10^-02
index	4.989*10^-03
N=5826 | R squared = 0.02954	

oilseed

v40	**Model 2.5**
	Coefficient (SE)
Intercept	5.291
gdp	7.206*10^-09
beds	5.011*10^-06
taps	-2.720*10^-02
index	-7.217*10^-02
N=11610 | R squared = 0.01624	

horticulture

v40	Model 2.6
	Coefficient (SE)
Intercept	4.932
gdp	7.449*10^-09
beds	8.159*10^-06
taps	-2.209*10^-02
index	-1.861*10^-02
N=12594 | R squared = 0.01806	

Q5)

The correlation coefficients of the explanatory variables are not equal to 0 and will not be totally independent.

As a result, the X matrix will not be full rank, the X transpose will be singular, and the transverse will not exist. As a result, model estimators would be obsolete.

Q6)

No, the relationship between yield increase and health indices varies by crop type.