Multivariate Analysis

libraries used:

```
library(tidyverse)
library(dplyr)
library("ggpubr")
library(factoextra)
```

Tidying Data

Checking for any missing values

```
## [1] 621
```

```
##                Site       Sampling_date   Water_temperature                  pH
##                   0                   0                  58                  47
##          Alkalinity    Suspended_solids          Phosphorus             Ammonium
##                  47                  49                  36                  80
##   Dissolved_silicon         Chlorophyll  Dissolved_fluoride  Dissolved_chloride
##                  25                  52                  25                  25
##   Dissolved_nitrate  Dissolved_sulphate    Dissolved_sodium Dissolved_potassium
##                  26                  25                  25                  26
##   Dissolved_calcium Dissolved_magnesium     Dissolved_boron
##                  25                  25                  25
```

Treating the missing values & looking at the Structure of the data.

The different samples or groups in the data set

```
##                                  Site
## 1              River Thame at Wheatley
## 204                 River Ray at Islip
## 407    River Cherwell at Hampton Poyle
## 610  River Evenlode at Cassington Mill
## 813           River Thames at Swinford
## 1016         River Thames at Newbridge
## 1219       River Windrush at Newbridge
## 1422 River Leach at Mill Lane,Lechlade
## 1625         River Cole at Lynt Bridge
## 1828            River Coln at Whelford
## 2031             River Ock at Abingdon
## 2234            River Pang at Tidmarsh
## 2437          River Thames at Sonning
## 2640          River Lodden at Charvil
## 2843          The Cut at Paley Street
```

```
## 3046         River Thames at Runnymede
## 3249           River Wye at Bourne End
## 3452       River Thames at Wallingford
## 3655   River Thames at Hannington Wick
## 3827       River Kennet at Woolhampton
## 3995         River Enborne at Brimpton
## 4163  Jubilee River at Pocock's Bridge
```

Checking if this is balanced design

```
design<-data %>% group_by(Site) %>%
  summarise(Count = n())
design
```

```
## # A tibble: 22 x 2
##    Site                              Count
##    <chr>                             <int>
##  1 Jubilee River at Pocock's Bridge    154
##  2 River Cherwell at Hampton Poyle     199
##  3 River Cole at Lynt Bridge           199
##  4 River Coln at Whelford              199
##  5 River Enborne at Brimpton           163
##  6 River Evenlode at Cassington Mill   199
##  7 River Kennet at Woolhampton         162
##  8 River Leach at Mill Lane,Lechlade   200
##  9 River Lodden at Charvil             195
## 10 River Ock at Abingdon               199
## # ... with 12 more rows
```

Since all the groups have almost equal number of observations we can assume this to be a balanced design.

Analysis of Data

MANOVA TEST

Null Hypothesis(Ho):There is no significant difference in mean vectors of all dependent variables. Alternate Hypothesis(H1):At least one mean vector that is significantly different.

```
manova1<-manova(dependentVar~data$Site)
##re<-data.frame(capture.output(summary(manova1)))
##knitr::kable(re, caption = "Manova output")
summary(manova1)
```

```
##             Df Pillai approx F num Df den Df    Pr(>F)
## data$Site   21 4.7874   77.952    357  70992 < 2.2e-16 ***
## Residuals 4176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova1,test = "Wilk")
```

```
##             Df     Wilks approx F num Df den Df    Pr(>F)
## data$Site   21 0.0001866   140.89    357  55210 < 2.2e-16 ***
## Residuals 4176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova1,test = "Roy")
```

```
##             Df    Roy approx F num Df den Df    Pr(>F)
## data$Site   21 11.779   2342.4     21   4176 < 2.2e-16 ***
## Residuals 4176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(manova1,test = "Hotelling-Lawley")
```

```
##             Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## data$Site   21           23.494   273.65    357  70688 < 2.2e-16 ***
## Residuals 4176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All P values are less than 0.05 hence the null hypothesis is rejected that is, at least one mean value of a dependent variable in a river is significantly different from others.

Individual means of all dependent variables with respect to its rivers

```
df1 = group_by(data, Site) %>%
  summarise_if(
    is.numeric, mean
  )

df1 = df1 %>% column_to_rownames(., var = "Site")##Making the column site just an index
df1
```

```
##                                 Water_temperature       pH Alkalinity
## Jubilee River at Pocock's Bridge          13.22273 7.951104   4088.130
## River Cherwell at Hampton Poyle           11.58492 7.910101   4133.769
## River Cole at Lynt Bridge                 11.83417 7.936633   4336.985
## River Coln at Whelford                    11.77990 7.987437   4245.558
## River Enborne at Brimpton                 10.54172 7.745767   2819.479
## River Evenlode at Cassington Mill         11.42010 7.908090   4023.588
## River Kennet at Woolhampton               11.31173 7.987407   4500.043
## River Leach at Mill Lane,Lechlade         11.43355 7.879650   4355.090
## River Lodden at Charvil                   12.25846 7.825077   3210.400
## River Ock at Abingdon                     11.71965 7.990251   4700.789
## River Pang at Tidmarsh                    10.87268 7.901443   4499.742
## River Ray at Islip                        11.38643 7.657839   4095.221
## River Thame at Wheatley                   11.67172 7.831566   4471.222
```

```
## River Thames at Hannington Wick            11.60238 7.881071   3955.810
## River Thames at Newbridge                  12.28492 7.974070   4162.593
## River Thames at Runnymede                  12.99439 7.935663   4000.673
## River Thames at Sonning                    12.34278 7.955876   4151.536
## River Thames at Swinford                   12.18241 7.996432   4065.226
## River Thames at Wallingford                12.76649 8.002423   4144.134
## River Windrush at Newbridge                11.59697 8.045556   3870.505
## River Wye at Bourne End                    12.27629 8.073454   4588.784
## The Cut at Paley Street                    12.59031 7.576224   2443.959
##                                  Suspended_solids Phosphorus   Ammonium
## Jubilee River at Pocock's Bridge         8.378312  191.57792 0.07147403
## River Cherwell at Hampton Poyle         13.327085  192.90452 0.03748744
## River Cole at Lynt Bridge               15.136884  306.84422 0.05323618
## River Coln at Whelford                   5.397437   83.50251 0.04296985
## River Enborne at Brimpton                9.513313  183.30675 0.07515337
## River Evenlode at Cassington Mill       15.569698  253.20603 0.04249246
## River Kennet at Woolhampton              9.280432   78.42593 0.04690123
## River Leach at Mill Lane,Lechlade        2.982100   34.32000 0.06053500
## River Lodden at Charvil                  7.257641  210.38974 0.07561538
## River Ock at Abingdon                   11.059246  320.28141 0.06213568
## River Pang at Tidmarsh                   8.240619   67.78866 0.03955155
## River Ray at Islip                       9.575729  515.44724 0.10907035
## River Thame at Wheatley                 14.012980  714.61111 0.23784343
## River Thames at Hannington Wick         11.673274  253.66667 0.09726786
## River Thames at Newbridge               10.972111  206.48744 0.05121106
## River Thames at Runnymede               11.808571  220.72449 0.08571939
## River Thames at Sonning                 11.474845  215.22680 0.07315979
## River Thames at Swinford                11.574673  171.58291 0.04721608
## River Thames at Wallingford             15.615412  300.63402 0.07719588
## River Windrush at Newbridge             13.928283  132.43939 0.04026768
## River Wye at Bourne End                 13.187680  287.13402 0.10841237
## The Cut at Paley Street                  9.218520  672.76020 0.21220918
##                                  Dissolved_silicon Chlorophyll
## Jubilee River at Pocock's Bridge          5.216948   18.656429
## River Cherwell at Hampton Poyle           3.269849   14.075729
## River Cole at Lynt Bridge                 6.414171    5.764523
## River Coln at Whelford                    2.599045    3.041055
## River Enborne at Brimpton                 6.876319    2.538834
## River Evenlode at Cassington Mill         2.681960   12.628995
## River Kennet at Woolhampton               6.809630    8.165309
## River Leach at Mill Lane,Lechlade         2.427750    1.958400
## River Lodden at Charvil                   5.437026    3.863692
## River Ock at Abingdon                     7.073216    3.887688
## River Pang at Tidmarsh                    6.950979    2.720979
## River Ray at Islip                        3.278291    7.979849
## River Thame at Wheatley                   6.537778   12.562525
## River Thames at Hannington Wick           3.530595    3.826488
## River Thames at Newbridge                 3.260050    9.661809
## River Thames at Runnymede                 4.989490   29.711122
## River Thames at Sonning                   5.035052   21.669175
## River Thames at Swinford                  2.948543   10.913266
## River Thames at Wallingford               4.199175   28.345361
## River Windrush at Newbridge               2.436364    4.108687
## River Wye at Bourne End                   6.701959    3.681031
```

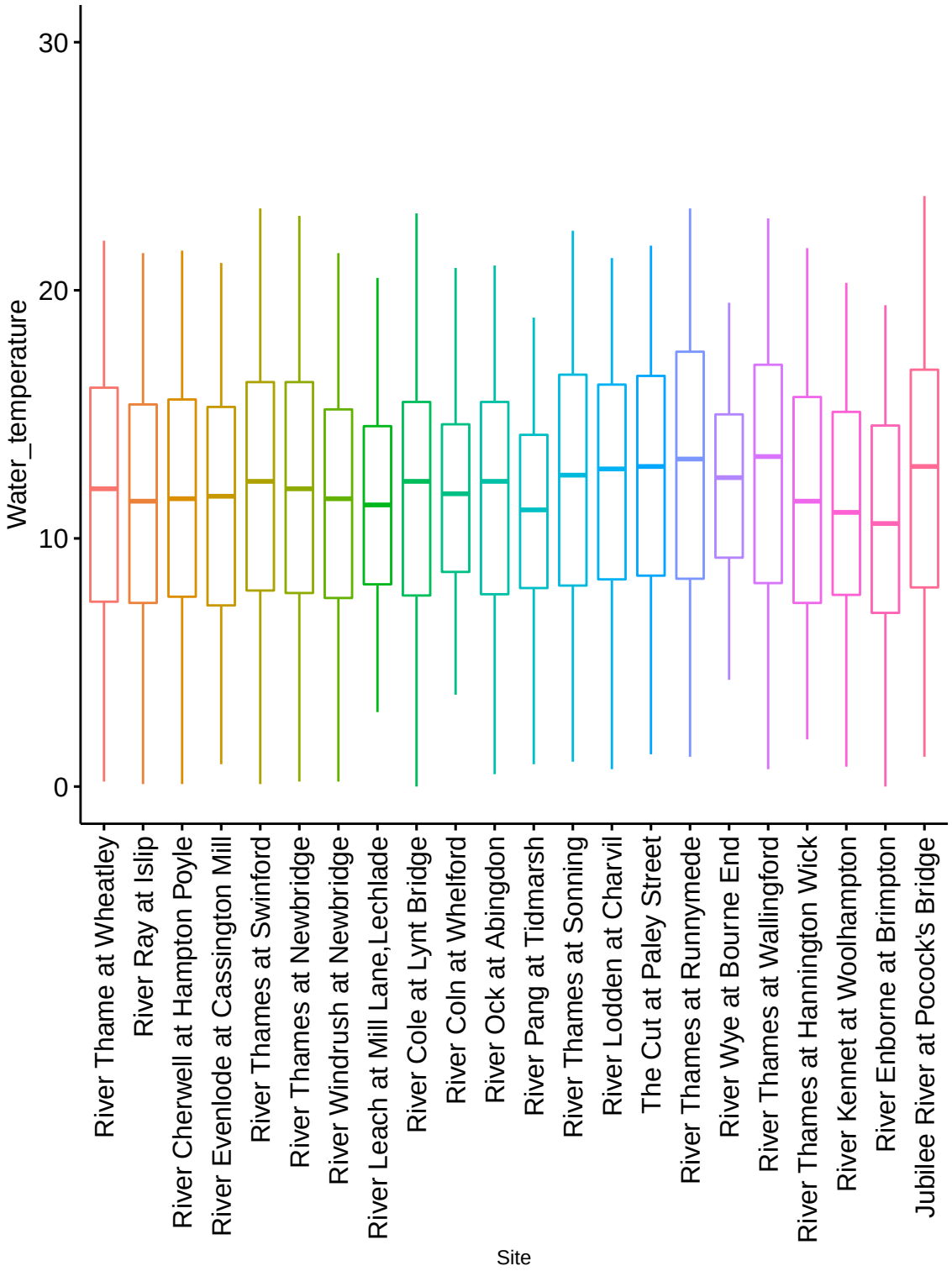
```
## The Cut at Paley Street                   6.211020    4.469235
##                                  Dissolved_fluoride Dissolved_chloride
## Jubilee River at Pocock's Bridge          0.1478571           43.74357
## River Cherwell at Hampton Poyle           0.2034171           53.91548
## River Cole at Lynt Bridge                 0.1935176           45.89060
## River Coln at Whelford                    0.1261809           16.57618
## River Enborne at Brimpton                 0.1180982           34.39706
## River Evenlode at Cassington Mill         0.1236683           25.65251
## River Kennet at Woolhampton               0.1204321           23.60309
## River Leach at Mill Lane,Lechlade         0.1034000           15.94700
## River Lodden at Charvil                   0.1212821           59.80564
## River Ock at Abingdon                     0.1961307           39.04121
## River Pang at Tidmarsh                    0.1357216           24.54665
## River Ray at Islip                        0.1805025           63.22116
## River Thame at Wheatley                   0.2160606           55.13985
## River Thames at Hannington Wick           0.1513095           56.30030
## River Thames at Newbridge                 0.1519598           40.48518
## River Thames at Runnymede                 0.1525000           45.89036
## River Thames at Sonning                   0.1556701           39.71624
## River Thames at Swinford                  0.1364824           35.25407
## River Thames at Wallingford               0.1688660           45.64758
## River Windrush at Newbridge               0.1048485           23.00737
## River Wye at Bourne End                   0.1056186           41.97938
## The Cut at Paley Street                   0.1691327           93.93398
##                                  Dissolved_nitrate Dissolved_sulphate
## Jubilee River at Pocock's Bridge          26.34688           47.30812
## River Cherwell at Hampton Poyle           25.13558           65.46955
## River Cole at Lynt Bridge                 18.38935           53.05261
## River Coln at Whelford                    26.20251           33.63749
## River Enborne at Brimpton                 17.16883           26.21638
## River Evenlode at Cassington Mill         24.62965           45.61819
## River Kennet at Woolhampton               24.03179           19.98204
## River Leach at Mill Lane,Lechlade         31.03035           35.13780
## River Lodden at Charvil                   34.36015           47.67841
## River Ock at Abingdon                     30.47839           71.65744
## River Pang at Tidmarsh                    28.00686           19.28082
## River Ray at Islip                        33.41749           94.39814
## River Thame at Wheatley                   35.02485           71.64040
## River Thames at Hannington Wick           32.30780           67.34893
## River Thames at Newbridge                 26.63367           53.54307
## River Thames at Runnymede                 28.00536           48.62082
## River Thames at Sonning                   26.62191           49.84732
## River Thames at Swinford                  25.94276           50.21307
## River Thames at Wallingford               28.09794           67.10814
## River Windrush at Newbridge               28.37136           42.03328
## River Wye at Bourne End                   27.49175           20.41031
## The Cut at Paley Street                   84.10010           99.50367
##                                  Dissolved_sodium Dissolved_potassium
## Jubilee River at Pocock's Bridge        27.404545            5.378571
## River Cherwell at Hampton Poyle         35.612563            6.214573
## River Cole at Lynt Bridge               27.381407            5.346231
## River Coln at Whelford                   8.777387            1.671859
## River Enborne at Brimpton               17.828221            3.575460
## River Evenlode at Cassington Mill       16.282412            3.561307
```

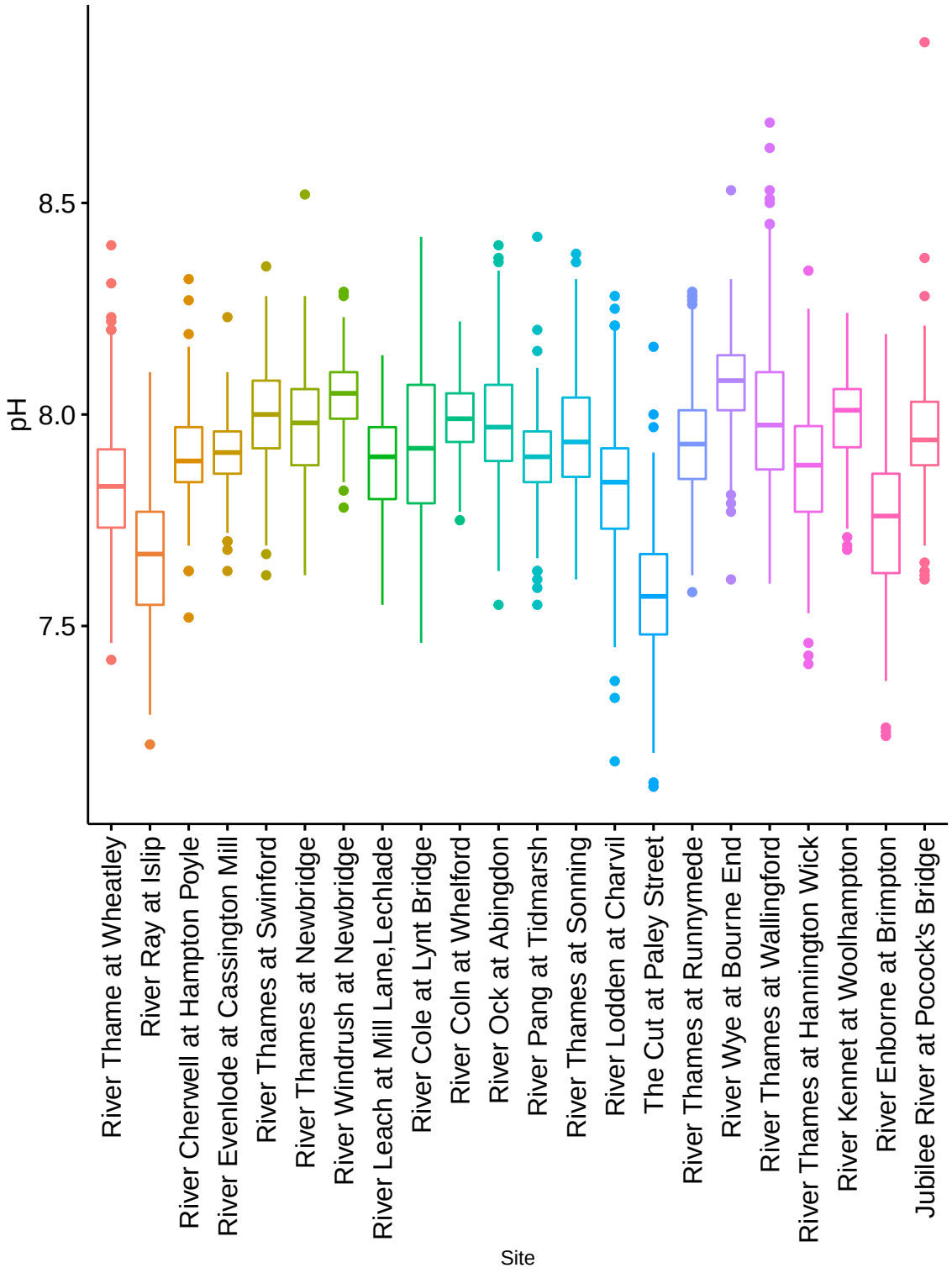
```
## River Kennet at Woolhampton              12.438272            2.375926
## River Leach at Mill Lane,Lechlade         8.376500            1.546000
## River Lodden at Charvil                  38.686667            7.515897
## River Ock at Abingdon                    25.012563            5.894975
## River Pang at Tidmarsh                   12.095876            2.882990
## River Ray at Islip                       48.773367           10.556784
## River Thame at Wheatley                  39.235354            9.655556
## River Thames at Hannington Wick          41.339286            8.138690
## River Thames at Newbridge                27.303518            5.488442
## River Thames at Runnymede                29.665816            5.846429
## River Thames at Sonning                  25.102062            5.205670
## River Thames at Swinford                 22.940201            4.622613
## River Thames at Wallingford              30.186598            6.431443
## River Windrush at Newbridge              13.392929            2.682828
## River Wye at Bourne End                  26.289175            4.231959
## The Cut at Paley Street                  70.796429           13.585714
##                                  Dissolved_calcium Dissolved_magnesium
## Jubilee River at Pocock's Bridge          101.68896             4.400649
## River Cherwell at Hampton Poyle           104.37889             7.568342
## River Cole at Lynt Bridge                 109.68693             4.371357
## River Coln at Whelford                    101.01256             5.747739
## River Enborne at Brimpton                  67.76626             4.370552
## River Evenlode at Cassington Mill         102.18291             4.220603
## River Kennet at Woolhampton               106.81852             2.171605
## River Leach at Mill Lane,Lechlade         109.39500             5.083000
## River Lodden at Charvil                    82.77538             5.233333
## River Ock at Abingdon                     126.37085             4.578392
## River Pang at Tidmarsh                    107.59536             3.206186
## River Ray at Islip                        112.20503             6.073869
## River Thame at Wheatley                   117.70808             5.407071
## River Thames at Hannington Wick           100.56369             5.206548
## River Thames at Newbridge                 103.57940             5.194975
## River Thames at Runnymede                 100.56071             4.570408
## River Thames at Sonning                   105.20515             4.440206
## River Thames at Swinford                  101.62915             4.973869
## River Thames at Wallingford               109.05155             5.357216
## River Windrush at Newbridge                98.37071             4.505051
## River Wye at Bourne End                   107.41753             1.918557
## The Cut at Paley Street                    84.51837             9.976531
##                                  Dissolved_boron
## Jubilee River at Pocock's Bridge        54.46948
## River Cherwell at Hampton Poyle         72.70754
## River Cole at Lynt Bridge               55.56533
## River Coln at Whelford                  19.90000
## River Enborne at Brimpton               25.90736
## River Evenlode at Cassington Mill       50.76131
## River Kennet at Woolhampton             21.50926
## River Leach at Mill Lane,Lechlade       24.56250
## River Lodden at Charvil                 56.66923
## River Ock at Abingdon                   62.17186
## River Pang at Tidmarsh                  20.71237
## River Ray at Islip                     107.68543
## River Thame at Wheatley                 86.72424
## River Thames at Hannington Wick         64.92917
```

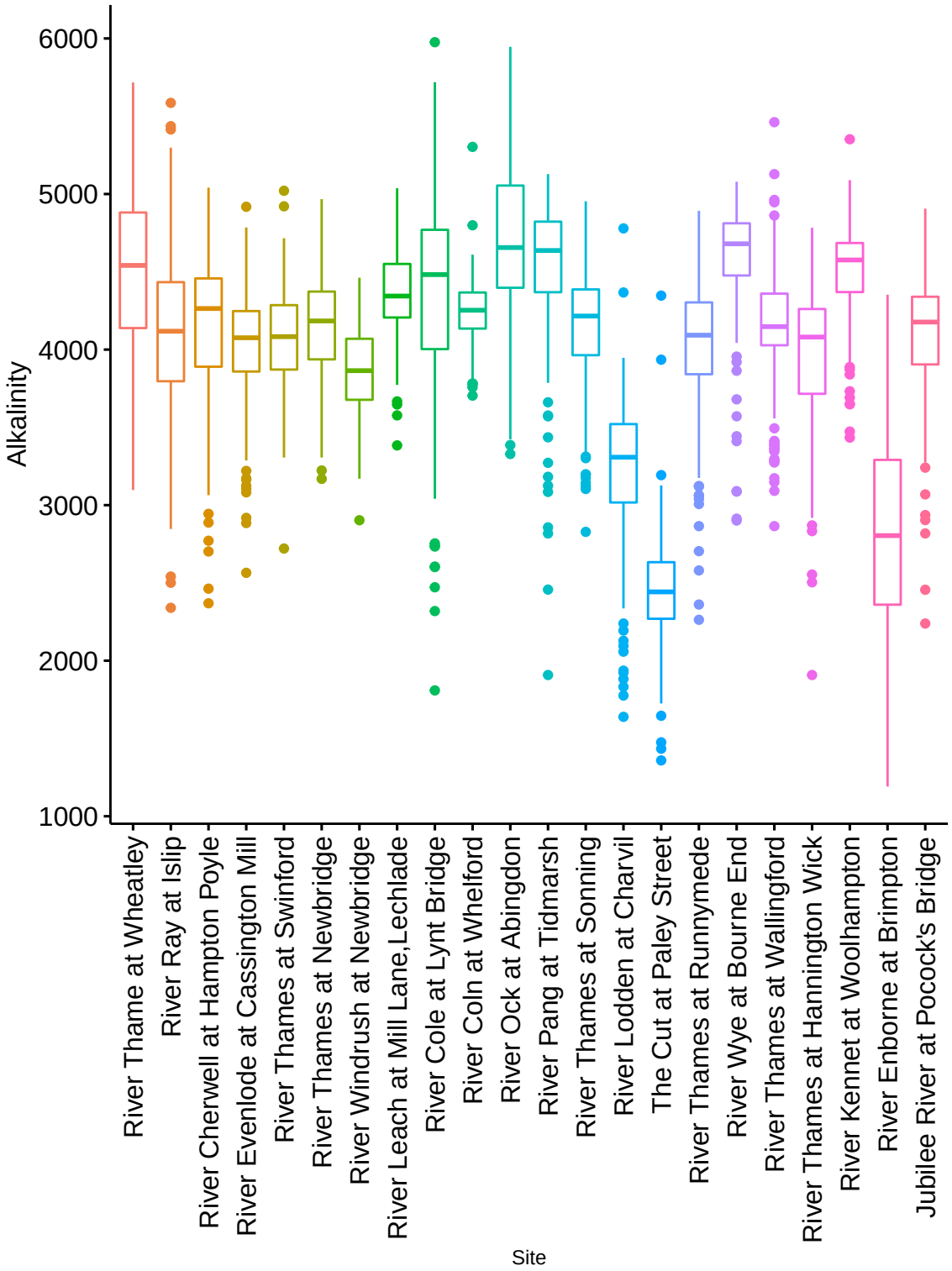
```
## River Thames at Newbridge               53.19648
## River Thames at Runnymede               60.53878
## River Thames at Sonning                 57.57371
## River Thames at Swinford                47.28492
## River Thames at Wallingford             76.31289
## River Windrush at Newbridge             32.96162
## River Wye at Bourne End                 35.09948
## The Cut at Paley Street                 89.49133
```

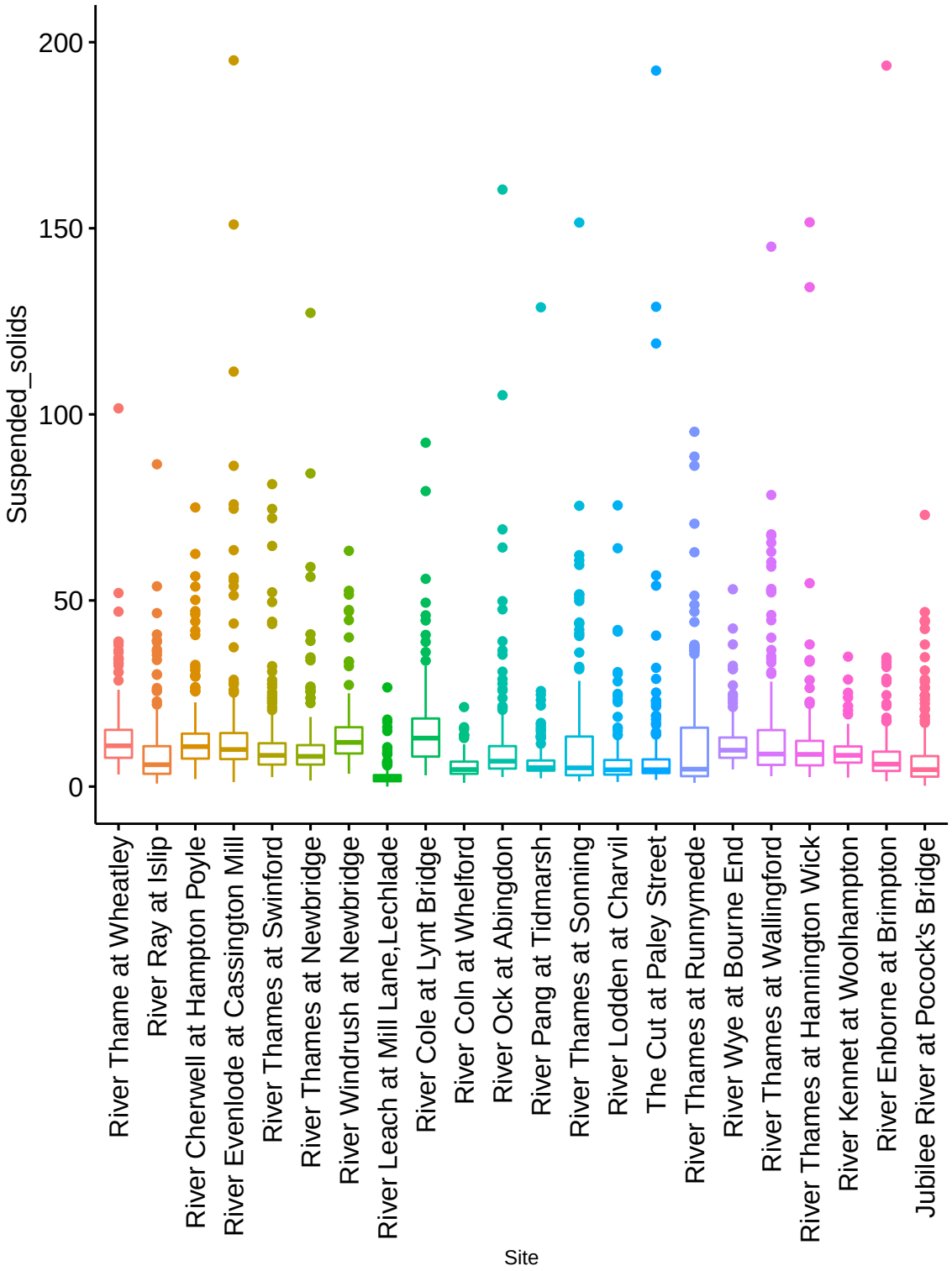
Visualization

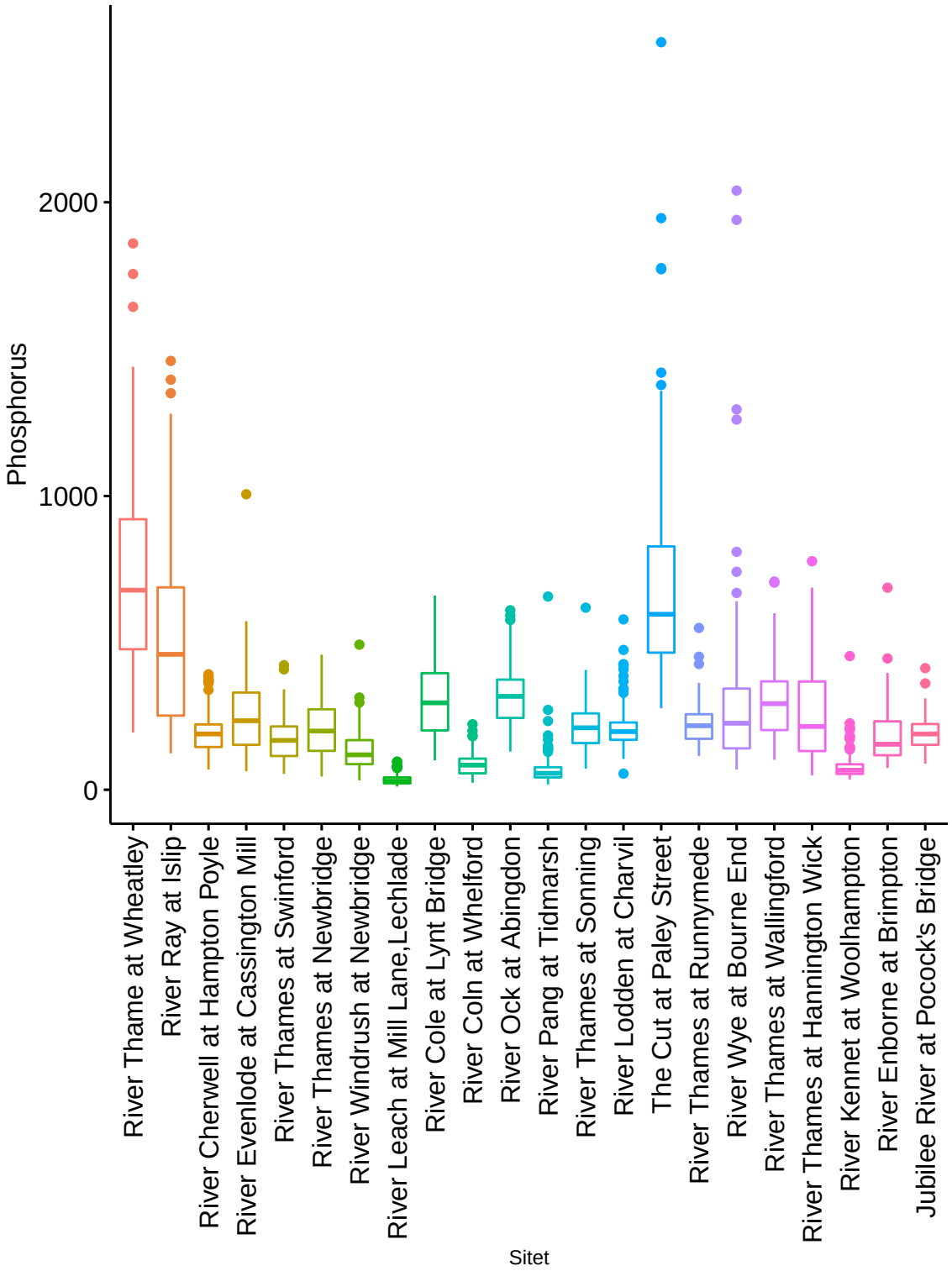
Let us Visualize how the first six dependent values of each river with respect to their sites.

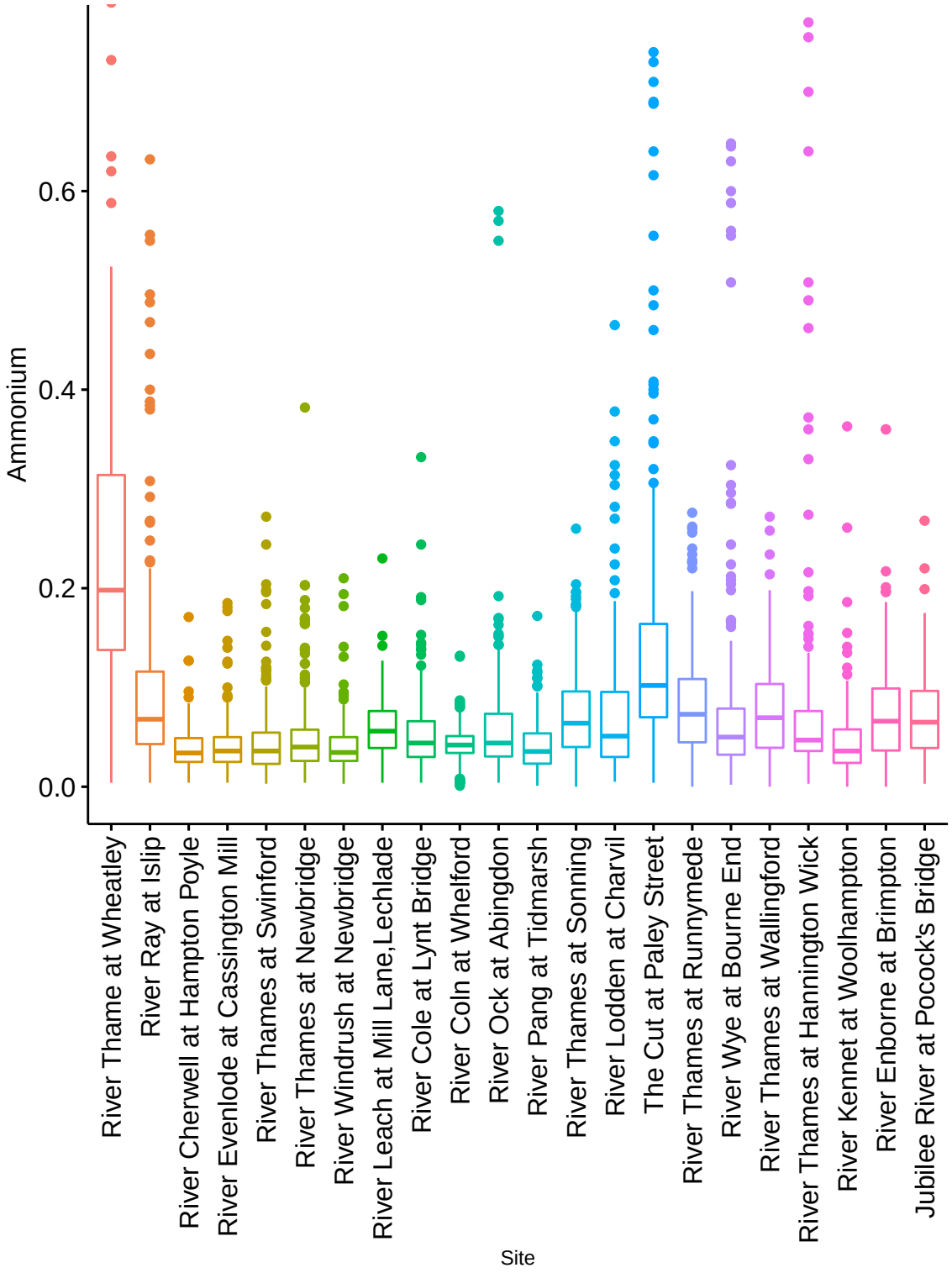

Water_temperature
30
20
10
0
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Site


pH
8.5
8.0
7.5
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Site


Alkalinity
6000
5000
4000
3000
2000
1000
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Site


Suspended_solids
200
150
100
50
0
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Site


Phosphorus
2000
1000
0
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Sitet


Ammonium
0.0
0.2
0.4
0.6
River Thame at Wheatley
River Ray at Islip
River Cherwell at Hampton Poyle
River Evenlode at Cassington Mill
River Thames at Swinford
River Thames at Newbridge
River Windrush at Newbridge
River Leach at Mill Lane,Lechlade
River Cole at Lynt Bridge
River Coln at Whelford
River Ock at Abingdon
River Pang at Tidmarsh
River Thames at Sonning
River Lodden at Charvil
The Cut at Paley Street
River Thames at Runnymede
River Wye at Bourne End
River Thames at Wallingford
River Thames at Hannington Wick
River Kennet at Woolhampton
River Enborne at Brimpton
Jubilee River at Pocock's Bridge
Site

We can see some dependent variables like pH and alkalinity has some significantly different values which confirms the rejection of null hypothesis.

Since the null hypothesis was rejected by the MANOVA test let us see the how the individual dependent variables behave by doing an ANOVA test for each response. Here response 1 is the water_temperature and the respective responses are in order as in the data set such that the response 17 is dissolved boron.

```
summary.aov(manova1)
```

```
##  Response 1 :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21   1656  78.848  3.1263 1.958e-06 ***
## Residuals   4176 105322  25.221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 2 :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 58.985 2.80879  140.21 < 2.2e-16 ***
## Residuals   4176 83.660 0.02003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 3 :
##               Df    Sum Sq  Mean Sq F value    Pr(>F)
## data$Site     21 1.197e+09 56998367  297.35 < 2.2e-16 ***
## Residuals   4176 8.005e+08   191691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 4 :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21  43713 2081.59  9.5908 < 2.2e-16 ***
## Residuals   4176 906360  217.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 5 :
##               Df    Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 126906003 6043143   270.6 < 2.2e-16 ***
## Residuals   4176  93260985   22333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 6 :
##               Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21  11.15 0.53093  15.168 < 2.2e-16 ***
## Residuals   4176 146.18 0.03500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 7 :
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 12103.8  576.37  350.69 < 2.2e-16 ***
```

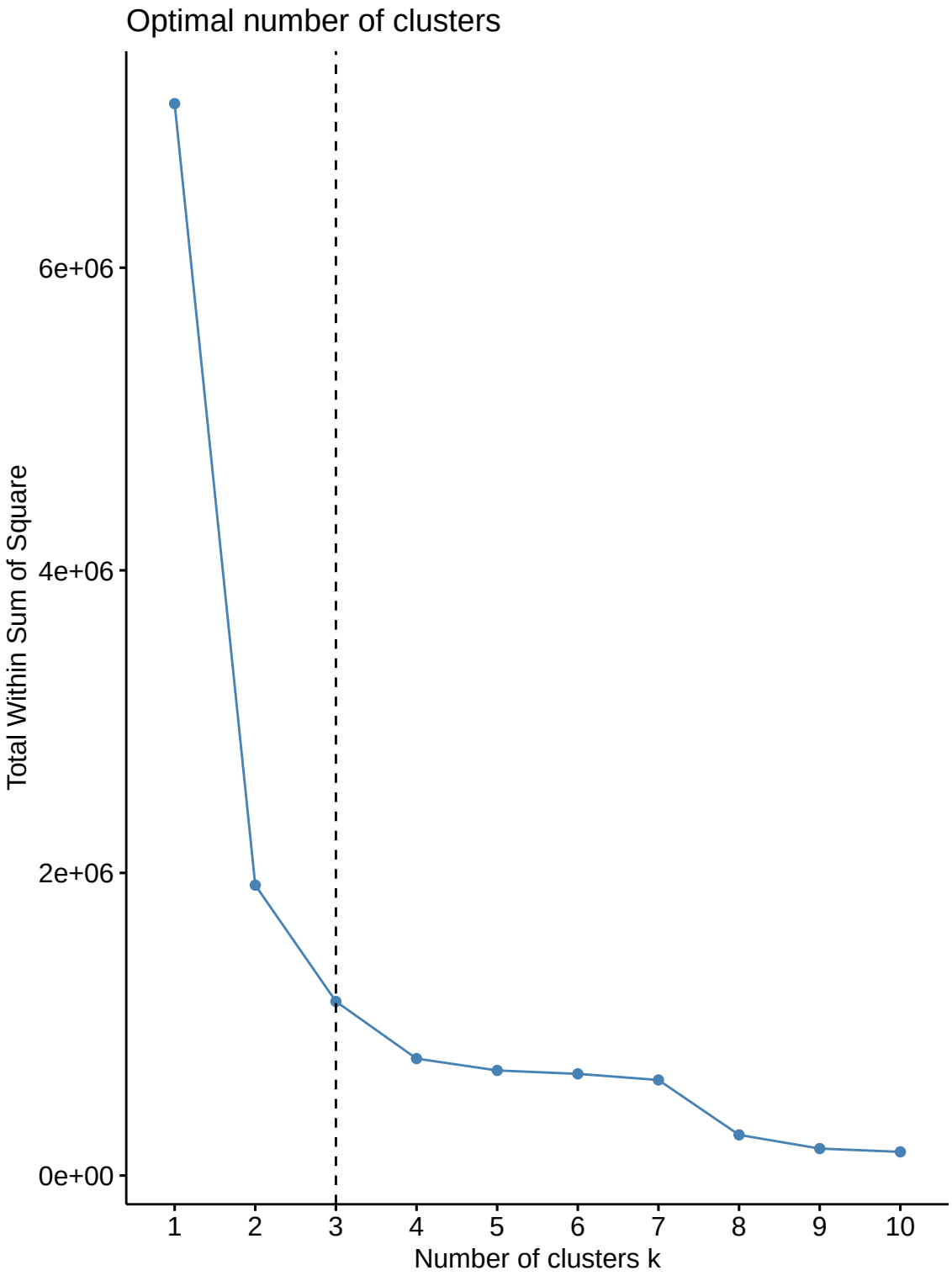
```
## Residuals   4176  6863.3    1.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 8 :
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21  273918 13043.7  17.591 < 2.2e-16 ***
## Residuals   4176 3096565   741.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 9 :
##              Df Sum Sq  Mean Sq F value    Pr(>F)
## data$Site     21 4.5824 0.218209  140.45 < 2.2e-16 ***
## Residuals   4176 6.4880 0.001554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 10 :
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 1310490   62404  315.85 < 2.2e-16 ***
## Residuals   4176  825082     198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 11 :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 672148   32007  377.14 < 2.2e-16 ***
## Residuals   4176 354407      85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 12 :
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 1912024   91049   557.3 < 2.2e-16 ***
## Residuals   4176  682256     163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 13 :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 868867   41375  329.14 < 2.2e-16 ***
## Residuals   4176 524941     126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 14 :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21  35942  1711.5  389.17 < 2.2e-16 ***
## Residuals   4176  18365     4.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 15 :
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 561217 26724.6  217.02 < 2.2e-16 ***
## Residuals   4176 514243   123.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 16 :
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 10850.8  516.71  794.73 < 2.2e-16 ***
## Residuals   4176  2715.1    0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 17 :
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## data$Site     21 2332235  111059  579.28 < 2.2e-16 ***
## Residuals   4176  800615     192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cluster Analysis

Finding the optimal number of clusters

```
fviz_nbclust(df1, kmeans, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
```



The optimal number of clusters is 3.

```
Cdata <- subset(data, select = -c(1,2,3))#Only have quantitatve variables

m <- apply(Cdata, 2, mean)
```

```r
str(Cdata)
```

```
## 'data.frame':    4198 obs. of  16 variables:
##  $ pH                 : num  8.01 7.94 8.05 8.14 8.2 8.2 8.11 8 7.98 7.9 ...
##  $ Alkalinity         : int  4915 5637 5393 5351 5129 5067 5076 4513 4088 4400 ...
##  $ Suspended_solids   : num  7.7 7.5 5.3 6 4.4 5.4 8.8 9.1 12 6.8 ...
##  $ Phosphorus         : int  438 341 415 381 480 568 568 747 1424 691 ...
##  $ Ammonium           : num  0.2 0.232 0.176 0.364 0.384 0.292 0.308 0.392 0.208 0.43 ...
##  $ Dissolved_silicon  : num  5.8 5.3 4.4 2.8 2.3 2.3 4.6 4.9 3.7 4.6 ...
##  $ Chlorophyll        : num  6.93 9.56 8.88 29.21 17.63 ...
##  $ Dissolved_fluoride : num  0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 ...
##  $ Dissolved_chloride : num  41 42.5 43.5 46 48.5 47.5 50 49 52 52 ...
##  $ Dissolved_nitrate  : num  34 30.5 30.5 36.5 34.5 35.5 34 34.5 37.5 34.5 ...
##  $ Dissolved_sulphate : num  77 81.5 80.5 76 70 68 69.5 71.5 73 75.5 ...
##  $ Dissolved_sodium   : num  26.7 29.7 29.4 34.5 36.9 34.2 39.3 34.4 40.2 41.7 ...
##  $ Dissolved_potassium: num  6.5 6.5 7.1 8 9 8.9 9.5 8.5 10 10.1 ...
##  $ Dissolved_calcium  : num  140 139 143 141 146 ...
##  $ Dissolved_magnesium: num  6 6.4 6.3 6.1 6.3 6.1 6.3 6.1 6 6.2 ...
##  $ Dissolved_boron    : num  81 88 89 83 79 91 94 96 104 107 ...
```

Non Hierarchical Clustering

```r
set.seed(123)
clusters = kmeans(df1,3)
#clusters$centers
df2 = df1
df2$Clusters <- as.factor(clusters$cluster)
clusters$cluster
```

```
##  Jubilee River at Pocock's Bridge   River Cherwell at Hampton Poyle
##                                 1                                  1
##         River Cole at Lynt Bridge            River Coln at Whelford
##                                 2                                  1
##         River Enborne at Brimpton River Evenlode at Cassington Mill
##                                 3                                  1
##       River Kennet at Woolhampton River Leach at Mill Lane,Lechlade
##                                 2                                  2
##           River Lodden at Charvil             River Ock at Abingdon
##                                 3                                  2
##            River Pang at Tidmarsh               River Ray at Islip
##                                 2                                  1
##           River Thame at Wheatley   River Thames at Hannington Wick
##                                 2                                  1
##         River Thames at Newbridge        River Thames at Runnymede
##                                 1                                  1
##           River Thames at Sonning         River Thames at Swinford
##                                 1                                  1
##       River Thames at Wallingford       River Windrush at Newbridge
##                                 1                                  1
##           River Wye at Bourne End           The Cut at Paley Street
##                                 2                                  3
```

```
df2
```

```
##                                  Water_temperature       pH Alkalinity
## Jubilee River at Pocock's Bridge          13.22273 7.951104   4088.130
## River Cherwell at Hampton Poyle           11.58492 7.910101   4133.769
## River Cole at Lynt Bridge                 11.83417 7.936633   4336.985
## River Coln at Whelford                    11.77990 7.987437   4245.558
## River Enborne at Brimpton                 10.54172 7.745767   2819.479
## River Evenlode at Cassington Mill         11.42010 7.908090   4023.588
## River Kennet at Woolhampton               11.31173 7.987407   4500.043
## River Leach at Mill Lane,Lechlade         11.43355 7.879650   4355.090
## River Lodden at Charvil                   12.25846 7.825077   3210.400
## River Ock at Abingdon                     11.71965 7.990251   4700.789
## River Pang at Tidmarsh                    10.87268 7.901443   4499.742
## River Ray at Islip                        11.38643 7.657839   4095.221
## River Thame at Wheatley                   11.67172 7.831566   4471.222
## River Thames at Hannington Wick           11.60238 7.881071   3955.810
## River Thames at Newbridge                 12.28492 7.974070   4162.593
## River Thames at Runnymede                 12.99439 7.935663   4000.673
## River Thames at Sonning                   12.34278 7.955876   4151.536
## River Thames at Swinford                  12.18241 7.996432   4065.226
## River Thames at Wallingford               12.76649 8.002423   4144.134
## River Windrush at Newbridge               11.59697 8.045556   3870.505
## River Wye at Bourne End                   12.27629 8.073454   4588.784
## The Cut at Paley Street                   12.59031 7.576224   2443.959
##                                  Suspended_solids Phosphorus   Ammonium
## Jubilee River at Pocock's Bridge         8.378312  191.57792 0.07147403
## River Cherwell at Hampton Poyle         13.327085  192.90452 0.03748744
## River Cole at Lynt Bridge               15.136884  306.84422 0.05323618
## River Coln at Whelford                   5.397437   83.50251 0.04296985
## River Enborne at Brimpton                9.513313  183.30675 0.07515337
## River Evenlode at Cassington Mill       15.569698  253.20603 0.04249246
## River Kennet at Woolhampton              9.280432   78.42593 0.04690123
## River Leach at Mill Lane,Lechlade        2.982100   34.32000 0.06053500
## River Lodden at Charvil                  7.257641  210.38974 0.07561538
## River Ock at Abingdon                   11.059246  320.28141 0.06213568
## River Pang at Tidmarsh                   8.240619   67.78866 0.03955155
## River Ray at Islip                       9.575729  515.44724 0.10907035
## River Thame at Wheatley                 14.012980  714.61111 0.23784343
## River Thames at Hannington Wick         11.673274  253.66667 0.09726786
## River Thames at Newbridge               10.972111  206.48744 0.05121106
## River Thames at Runnymede               11.808571  220.72449 0.08571939
## River Thames at Sonning                 11.474845  215.22680 0.07315979
## River Thames at Swinford                11.574673  171.58291 0.04721608
## River Thames at Wallingford             15.615412  300.63402 0.07719588
## River Windrush at Newbridge             13.928283  132.43939 0.04026768
## River Wye at Bourne End                 13.187680  287.13402 0.10841237
## The Cut at Paley Street                  9.218520  672.76020 0.21220918
##                                  Dissolved_silicon Chlorophyll
## Jubilee River at Pocock's Bridge          5.216948   18.656429
## River Cherwell at Hampton Poyle           3.269849   14.075729
## River Cole at Lynt Bridge                 6.414171    5.764523
## River Coln at Whelford                    2.599045    3.041055
## River Enborne at Brimpton                 6.876319    2.538834
```

```
## River Evenlode at Cassington Mill          2.681960   12.628995
## River Kennet at Woolhampton                6.809630    8.165309
## River Leach at Mill Lane,Lechlade          2.427750    1.958400
## River Lodden at Charvil                    5.437026    3.863692
## River Ock at Abingdon                      7.073216    3.887688
## River Pang at Tidmarsh                     6.950979    2.720979
## River Ray at Islip                         3.278291    7.979849
## River Thame at Wheatley                    6.537778   12.562525
## River Thames at Hannington Wick            3.530595    3.826488
## River Thames at Newbridge                  3.260050    9.661809
## River Thames at Runnymede                  4.989490   29.711122
## River Thames at Sonning                    5.035052   21.669175
## River Thames at Swinford                   2.948543   10.913266
## River Thames at Wallingford                4.199175   28.345361
## River Windrush at Newbridge                2.436364    4.108687
## River Wye at Bourne End                    6.701959    3.681031
## The Cut at Paley Street                    6.211020    4.469235
##                                  Dissolved_fluoride Dissolved_chloride
## Jubilee River at Pocock's Bridge          0.1478571           43.74357
## River Cherwell at Hampton Poyle           0.2034171           53.91548
## River Cole at Lynt Bridge                 0.1935176           45.89060
## River Coln at Whelford                    0.1261809           16.57618
## River Enborne at Brimpton                 0.1180982           34.39706
## River Evenlode at Cassington Mill         0.1236683           25.65251
## River Kennet at Woolhampton               0.1204321           23.60309
## River Leach at Mill Lane,Lechlade         0.1034000           15.94700
## River Lodden at Charvil                   0.1212821           59.80564
## River Ock at Abingdon                     0.1961307           39.04121
## River Pang at Tidmarsh                    0.1357216           24.54665
## River Ray at Islip                        0.1805025           63.22116
## River Thame at Wheatley                   0.2160606           55.13985
## River Thames at Hannington Wick           0.1513095           56.30030
## River Thames at Newbridge                 0.1519598           40.48518
## River Thames at Runnymede                 0.1525000           45.89036
## River Thames at Sonning                   0.1556701           39.71624
## River Thames at Swinford                  0.1364824           35.25407
## River Thames at Wallingford               0.1688660           45.64758
## River Windrush at Newbridge               0.1048485           23.00737
## River Wye at Bourne End                   0.1056186           41.97938
## The Cut at Paley Street                   0.1691327           93.93398
##                                  Dissolved_nitrate Dissolved_sulphate
## Jubilee River at Pocock's Bridge          26.34688           47.30812
## River Cherwell at Hampton Poyle           25.13558           65.46955
## River Cole at Lynt Bridge                 18.38935           53.05261
## River Coln at Whelford                    26.20251           33.63749
## River Enborne at Brimpton                 17.16883           26.21638
## River Evenlode at Cassington Mill         24.62965           45.61819
## River Kennet at Woolhampton               24.03179           19.98204
## River Leach at Mill Lane,Lechlade         31.03035           35.13780
## River Lodden at Charvil                   34.36015           47.67841
## River Ock at Abingdon                     30.47839           71.65744
## River Pang at Tidmarsh                    28.00686           19.28082
## River Ray at Islip                        33.41749           94.39814
## River Thame at Wheatley                   35.02485           71.64040
```

```
## River Thames at Hannington Wick            32.30780           67.34893
## River Thames at Newbridge                  26.63367           53.54307
## River Thames at Runnymede                  28.00536           48.62082
## River Thames at Sonning                    26.62191           49.84732
## River Thames at Swinford                   25.94276           50.21307
## River Thames at Wallingford                28.09794           67.10814
## River Windrush at Newbridge                28.37136           42.03328
## River Wye at Bourne End                    27.49175           20.41031
## The Cut at Paley Street                    84.10010           99.50367
##                                  Dissolved_sodium Dissolved_potassium
## Jubilee River at Pocock's Bridge        27.404545            5.378571
## River Cherwell at Hampton Poyle         35.612563            6.214573
## River Cole at Lynt Bridge               27.381407            5.346231
## River Coln at Whelford                   8.777387            1.671859
## River Enborne at Brimpton               17.828221            3.575460
## River Evenlode at Cassington Mill       16.282412            3.561307
## River Kennet at Woolhampton             12.438272            2.375926
## River Leach at Mill Lane,Lechlade        8.376500            1.546000
## River Lodden at Charvil                 38.686667            7.515897
## River Ock at Abingdon                   25.012563            5.894975
## River Pang at Tidmarsh                  12.095876            2.882990
## River Ray at Islip                      48.773367           10.556784
## River Thame at Wheatley                 39.235354            9.655556
## River Thames at Hannington Wick         41.339286            8.138690
## River Thames at Newbridge               27.303518            5.488442
## River Thames at Runnymede               29.665816            5.846429
## River Thames at Sonning                 25.102062            5.205670
## River Thames at Swinford                22.940201            4.622613
## River Thames at Wallingford             30.186598            6.431443
## River Windrush at Newbridge             13.392929            2.682828
## River Wye at Bourne End                 26.289175            4.231959
## The Cut at Paley Street                 70.796429           13.585714
##                                  Dissolved_calcium Dissolved_magnesium
## Jubilee River at Pocock's Bridge         101.68896            4.400649
## River Cherwell at Hampton Poyle          104.37889            7.568342
## River Cole at Lynt Bridge                109.68693            4.371357
## River Coln at Whelford                   101.01256            5.747739
## River Enborne at Brimpton                 67.76626            4.370552
## River Evenlode at Cassington Mill        102.18291            4.220603
## River Kennet at Woolhampton              106.81852            2.171605
## River Leach at Mill Lane,Lechlade        109.39500            5.083000
## River Lodden at Charvil                   82.77538            5.233333
## River Ock at Abingdon                    126.37085            4.578392
## River Pang at Tidmarsh                   107.59536            3.206186
## River Ray at Islip                       112.20503            6.073869
## River Thame at Wheatley                  117.70808            5.407071
## River Thames at Hannington Wick          100.56369            5.206548
## River Thames at Newbridge                103.57940            5.194975
## River Thames at Runnymede                100.56071            4.570408
## River Thames at Sonning                  105.20515            4.440206
## River Thames at Swinford                 101.62915            4.973869
## River Thames at Wallingford              109.05155            5.357216
## River Windrush at Newbridge               98.37071            4.505051
## River Wye at Bourne End                  107.41753            1.918557
```

```
## The Cut at Paley Street                    84.51837           9.976531
##                                  Dissolved_boron Clusters
## Jubilee River at Pocock's Bridge         54.46948        1
## River Cherwell at Hampton Poyle          72.70754        1
## River Cole at Lynt Bridge                55.56533        2
## River Coln at Whelford                   19.90000        1
## River Enborne at Brimpton                25.90736        3
## River Evenlode at Cassington Mill        50.76131        1
## River Kennet at Woolhampton              21.50926        2
## River Leach at Mill Lane,Lechlade        24.56250        2
## River Lodden at Charvil                  56.66923        3
## River Ock at Abingdon                    62.17186        2
## River Pang at Tidmarsh                   20.71237        2
## River Ray at Islip                      107.68543        1
## River Thame at Wheatley                  86.72424        2
## River Thames at Hannington Wick          64.92917        1
## River Thames at Newbridge                53.19648        1
## River Thames at Runnymede                60.53878        1
## River Thames at Sonning                  57.57371        1
## River Thames at Swinford                 47.28492        1
## River Thames at Wallingford              76.31289        1
## River Windrush at Newbridge              32.96162        1
## River Wye at Bourne End                  35.09948        2
## The Cut at Paley Street                  89.49133        3
```

Interpreting the clusters

```
dff = aggregate(df1, by=list(cluster=clusters$cluster), mean)#Means for each dependent variable in clus
dff
```

```
##   cluster Water_temperature       pH Alkalinity Suspended_solids Phosphorus
## 1       1          12.09704 7.933805   4078.062        11.607953   228.1167
## 2       2          11.58854 7.942915   4493.236        10.557135   258.4865
## 3       3          11.79683 7.715689   2824.613         8.663158   355.4856
##     Ammonium Dissolved_silicon Chlorophyll Dissolved_fluoride
## 1 0.06462765          3.620447   13.718164          0.1502719
## 2 0.08694506          6.130783    5.534351          0.1529830
## 3 0.12099265          6.174788    3.623920          0.1361710
##   Dissolved_chloride Dissolved_nitrate Dissolved_sulphate Dissolved_sodium
## 1           40.78417          27.64274           55.42884         27.23172
## 2           35.16397          27.77905           41.59449         21.54702
## 3           62.71223          45.20970           57.79949         42.43711
##   Dissolved_potassium Dissolved_calcium Dissolved_magnesium Dissolved_boron
## 1            5.483268         103.36906            5.188290        58.19344
## 2            4.561948         112.14175            3.819452        43.76358
## 3            8.225691          78.35334            6.526805        57.35597
```

```
##knitr::kable(dff,digits = 3)
```

By Clustering we have grouped the river sites according to similarity measures and we have obtained a summary of the clusters, from this we can see the differences in each cluster clearly.

```
#sites<- c(unique(data["Site"]))

#library(ggmap)
#register_google(key = "...")
#lonlat <- geocode(location = order1)
```