

Hw2

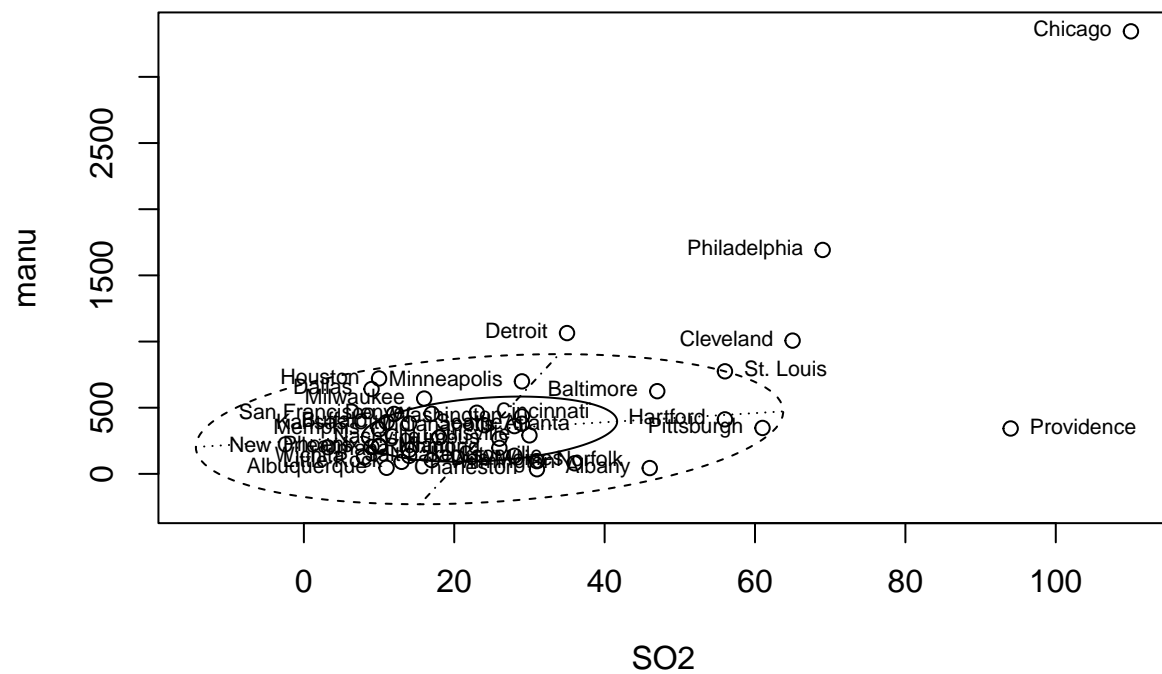
Getong Zhong

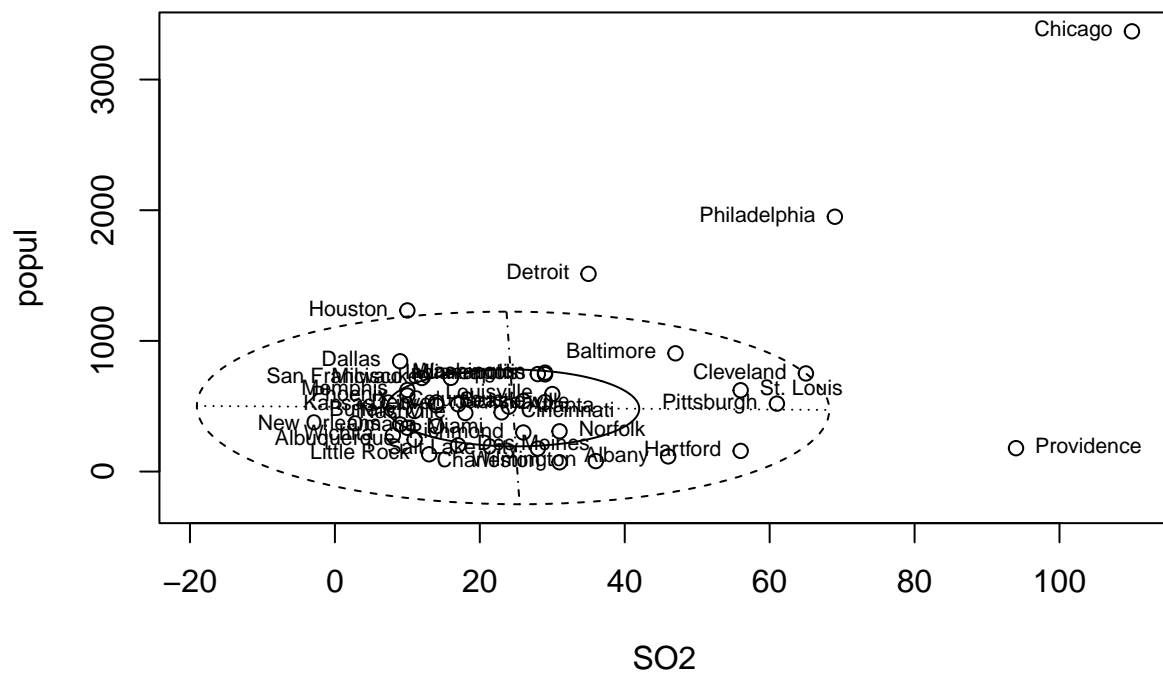
2023-02-12

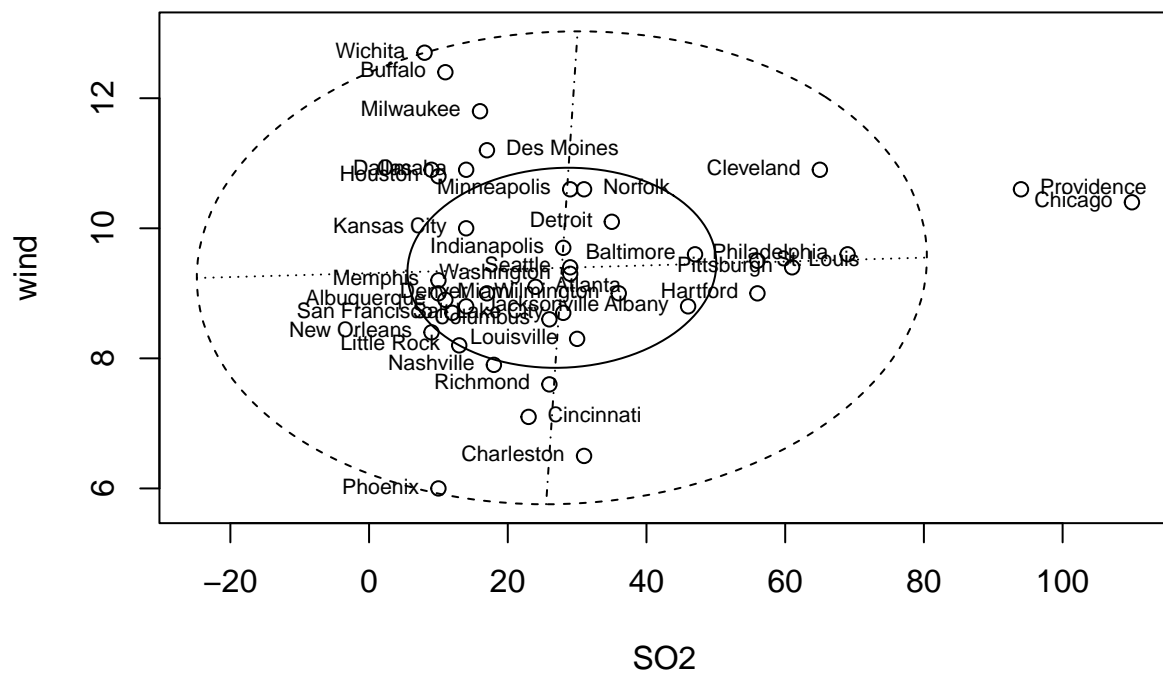
2.1

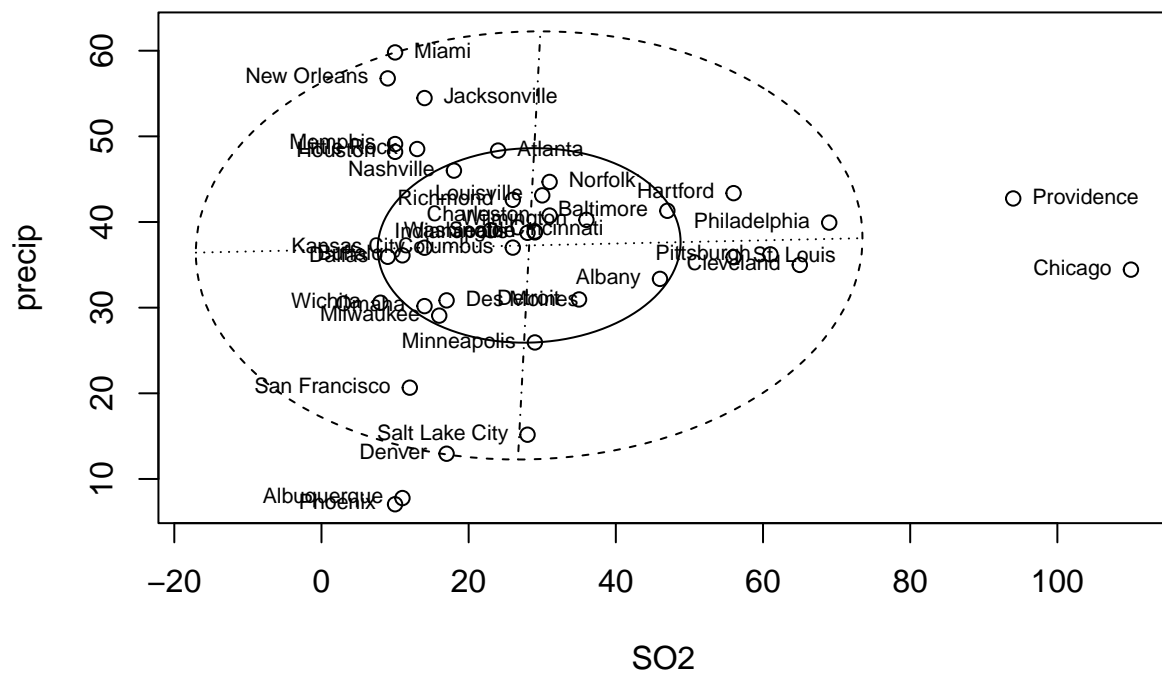
```
data("USairpollution", package = "HSAUR2")

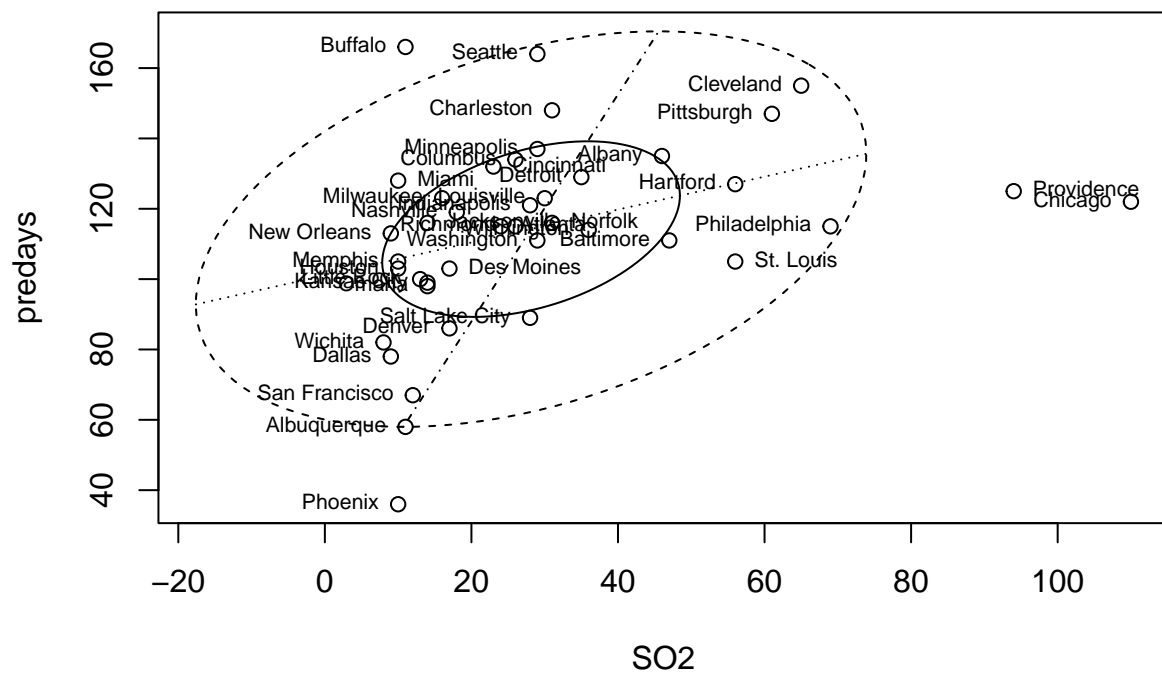
for(i in seq(1, 7)){
  for(j in seq(i, 7)) {
    if (i != j) {
      bvxbox(USairpollution[, c(i, j)], xlab=names(USairpollution)[i], ylab=names(USairpollution)[j])
      text(USairpollution[, c(i, j)][,1], USairpollution[, c(i, j)][,2], labels = rownames(USairpollution)[c(i, j)])
      cor(USairpollution[, c(i, j)][,1], USairpollution[, c(i, j)][,2])
    }
  }
}
```

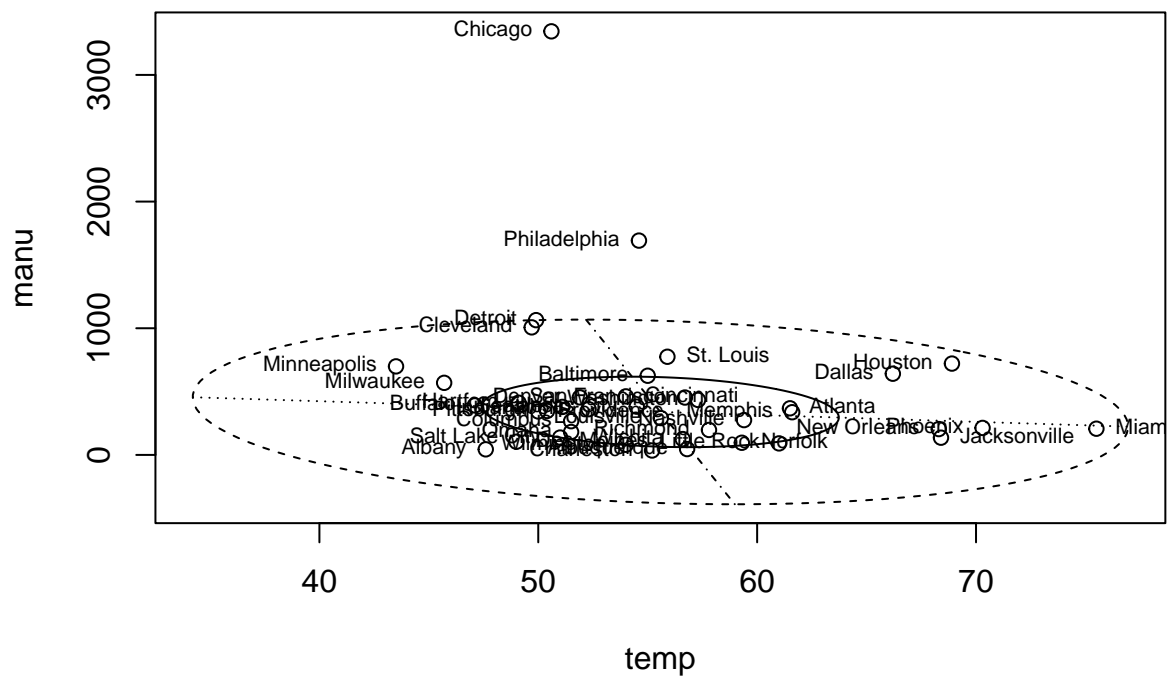



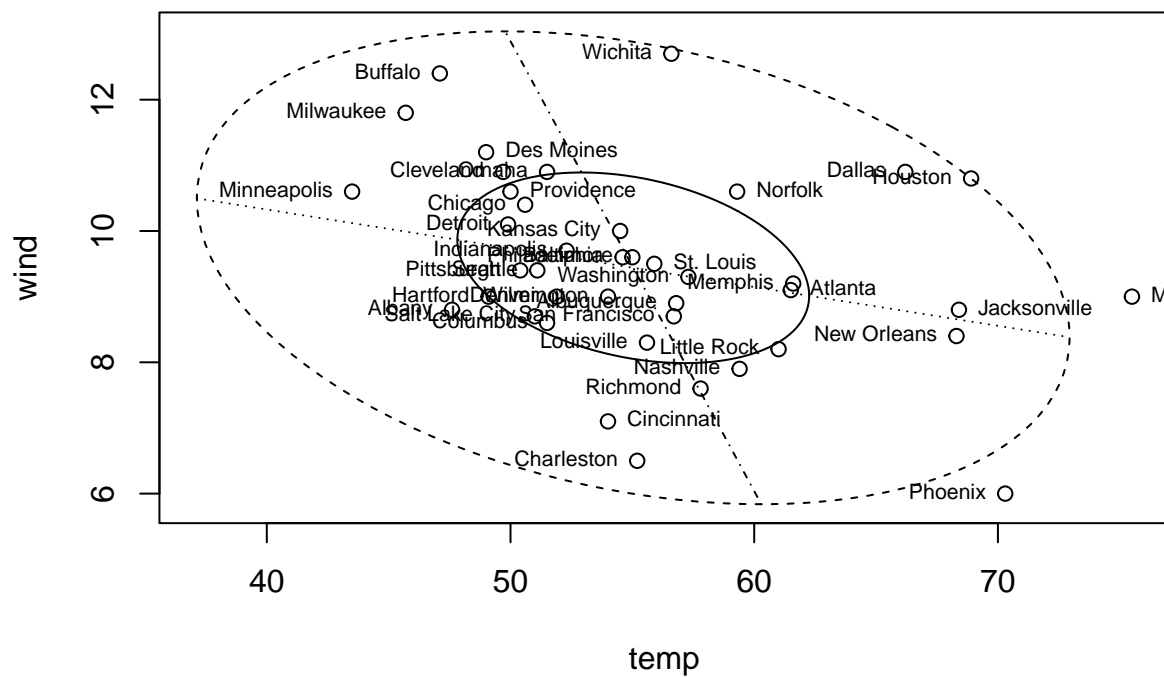


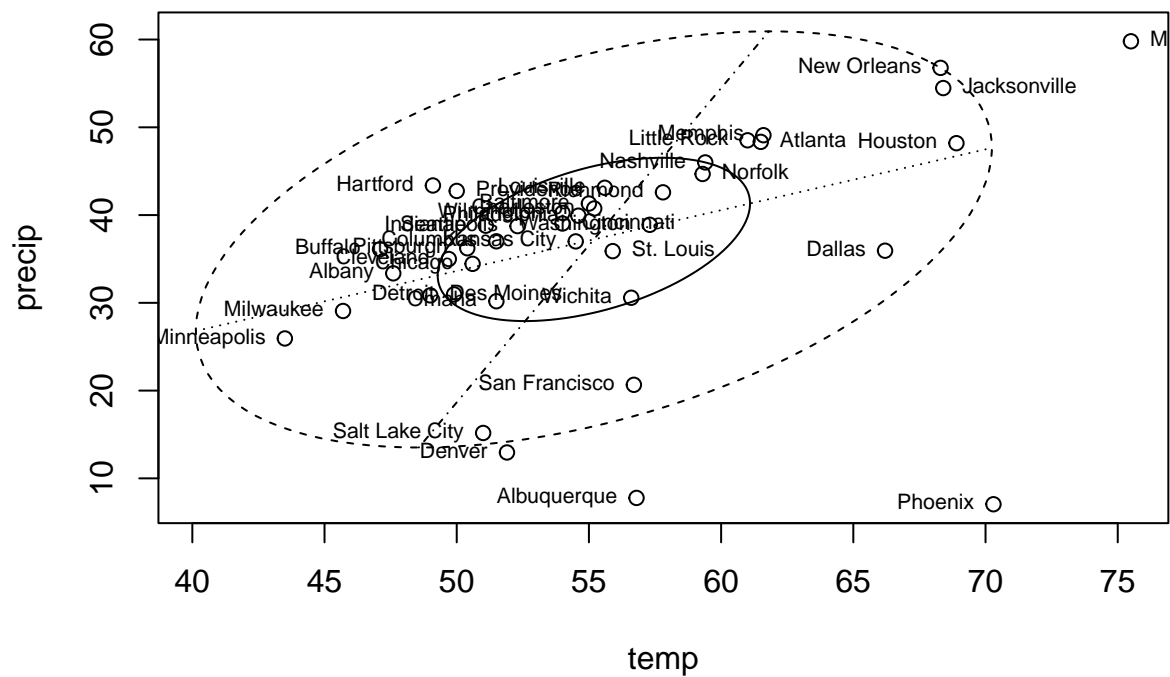


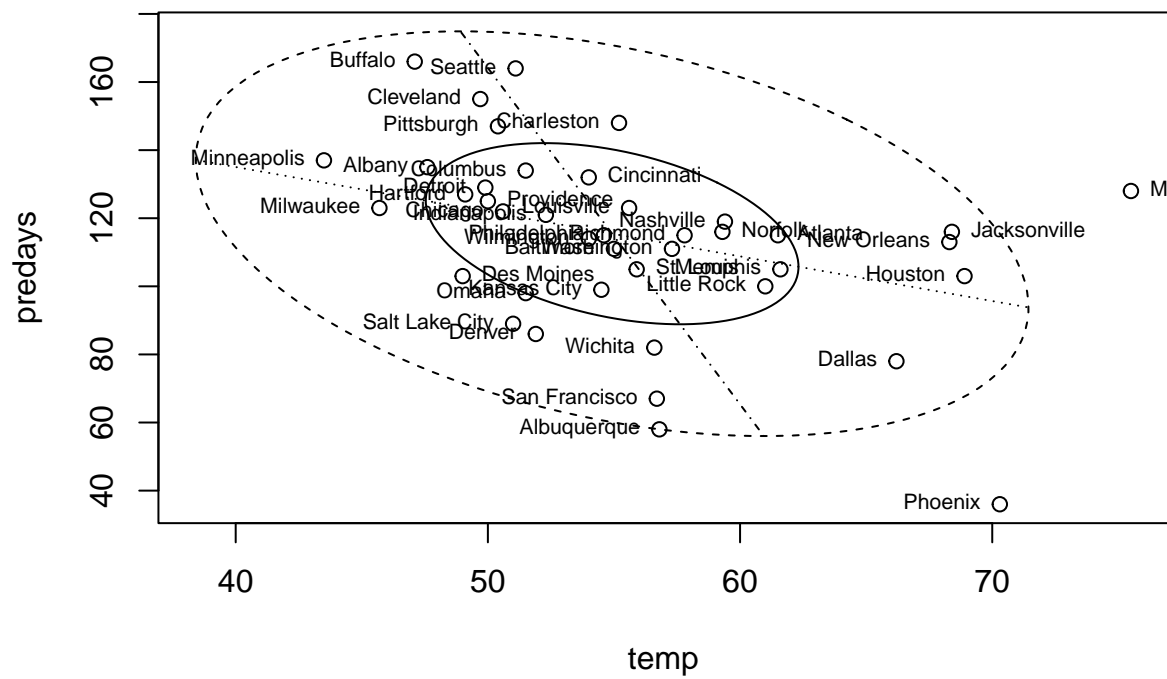


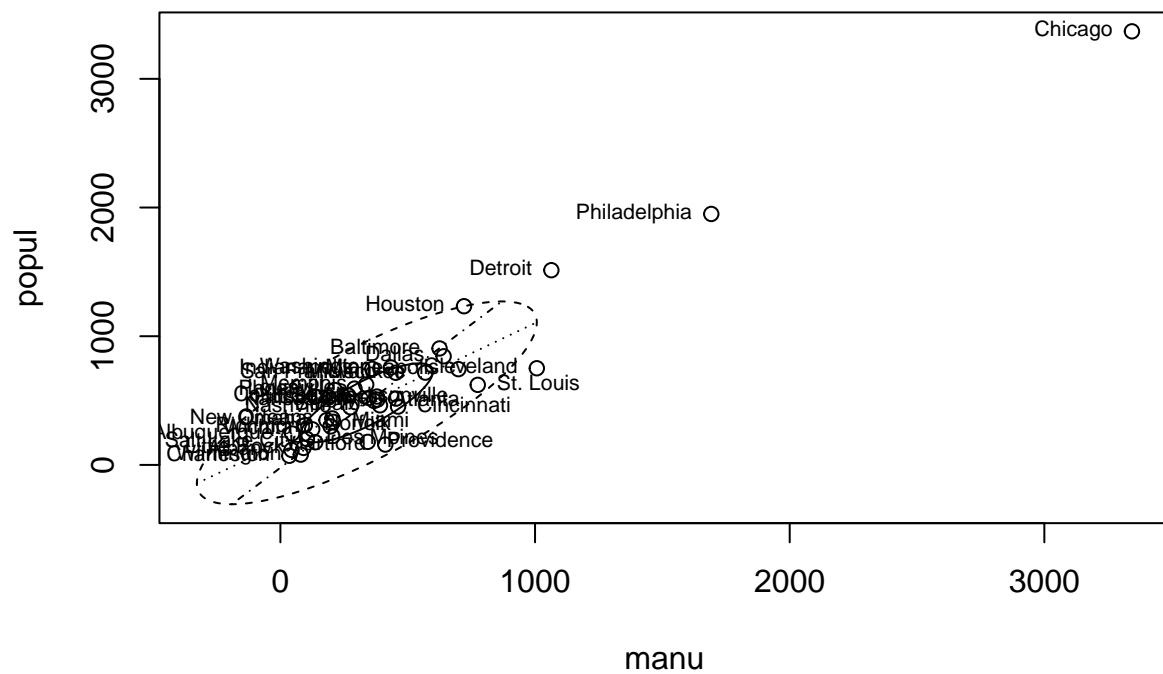


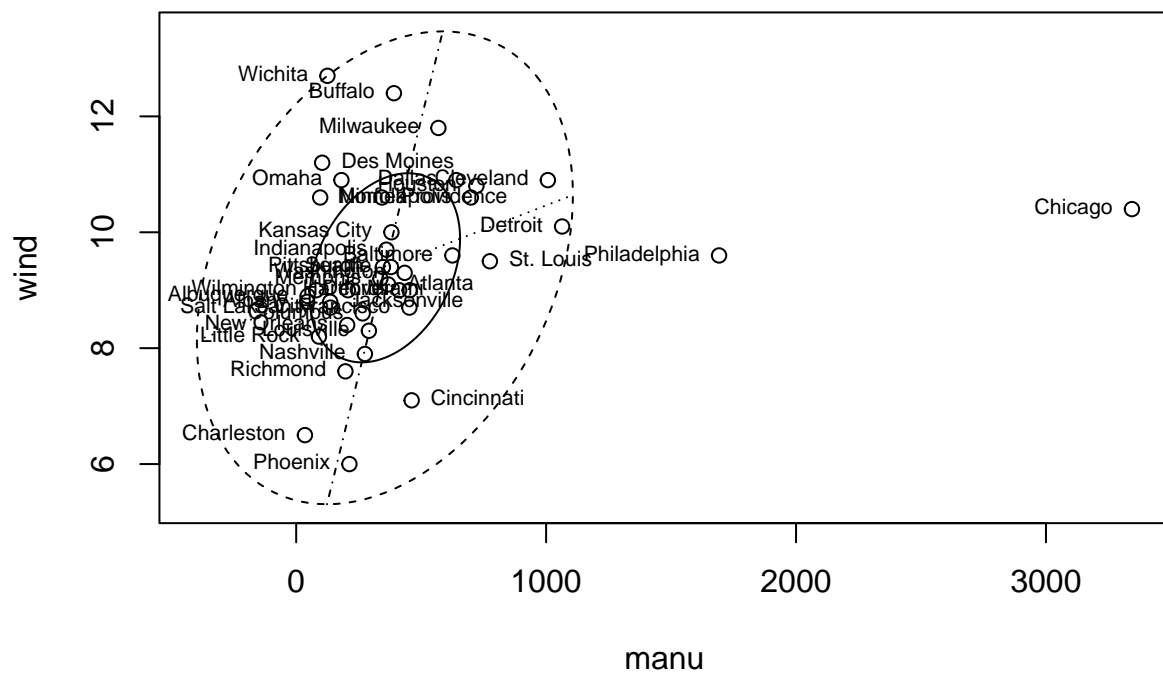


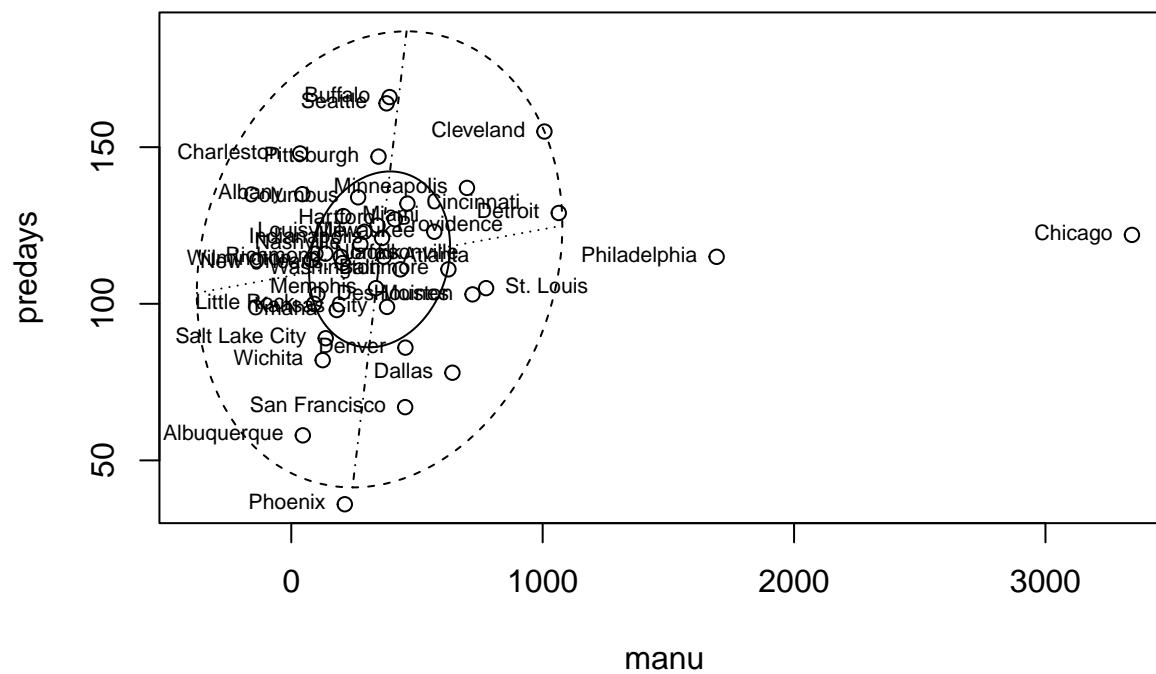


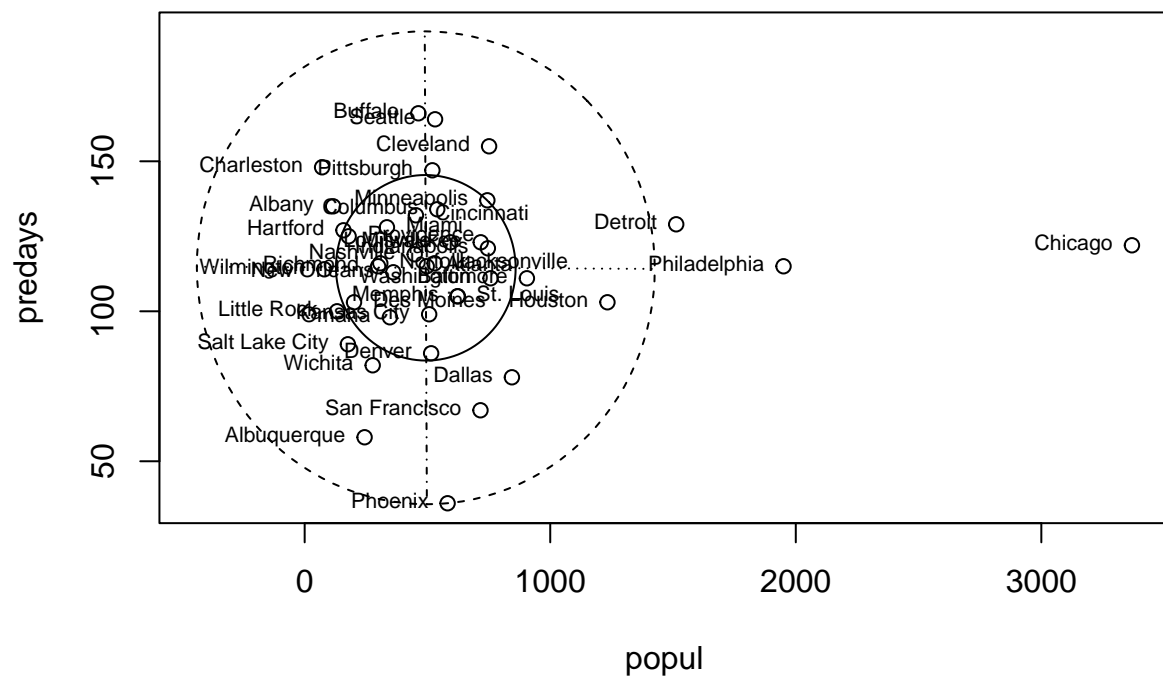


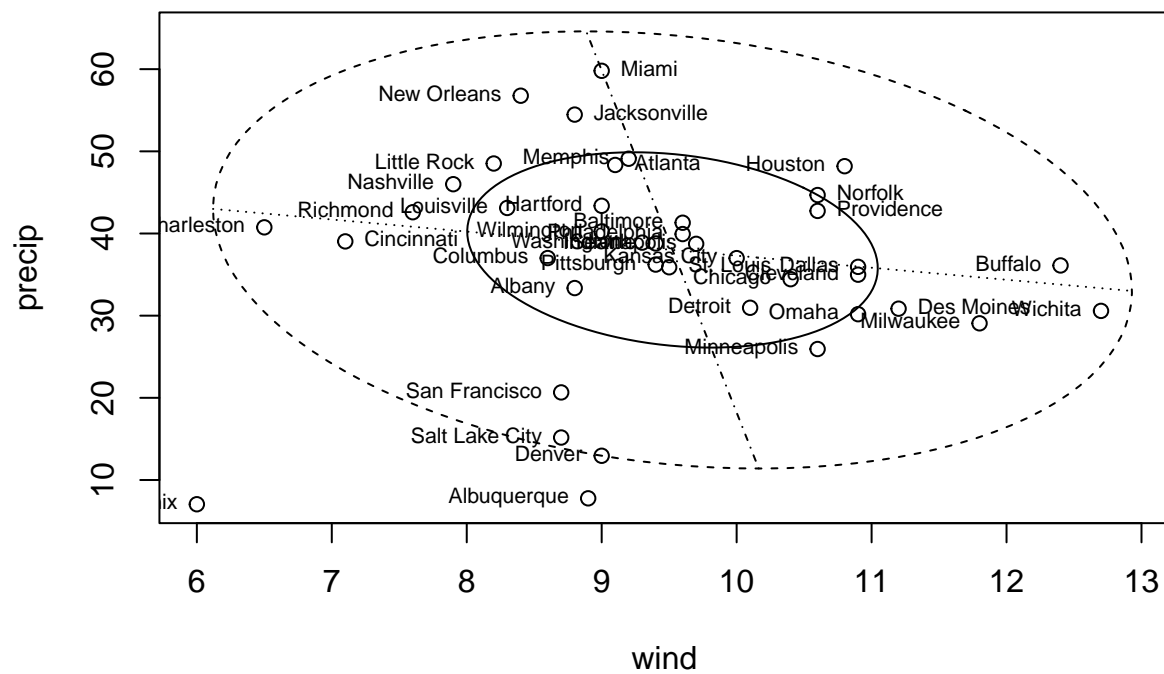


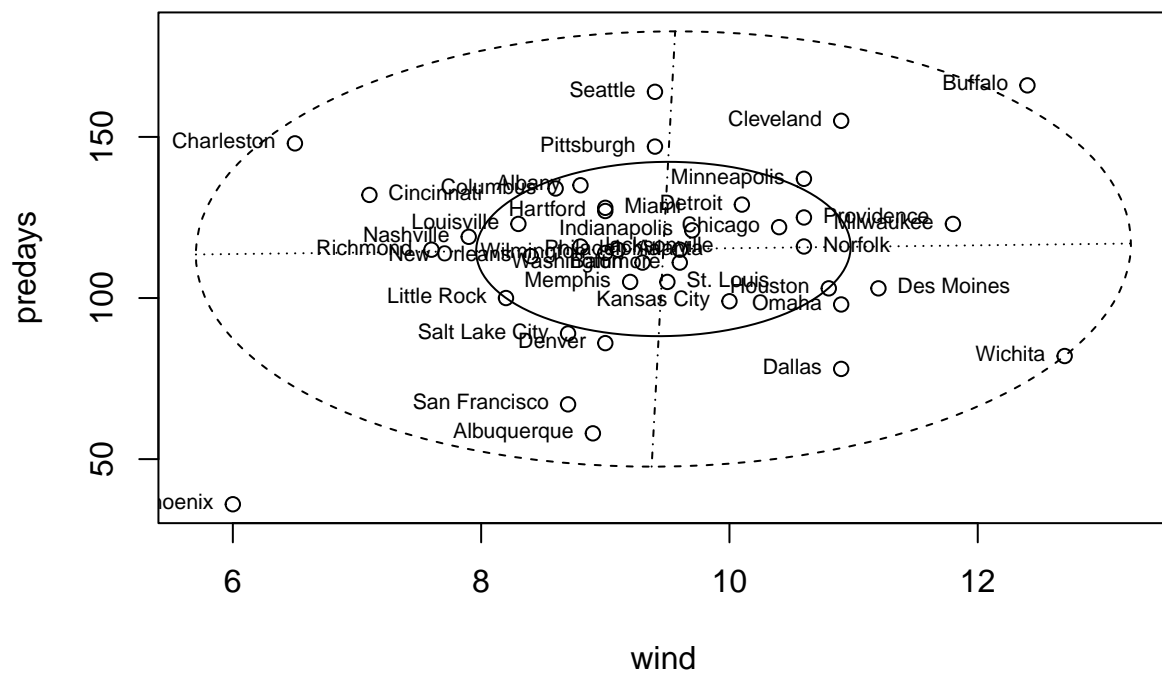


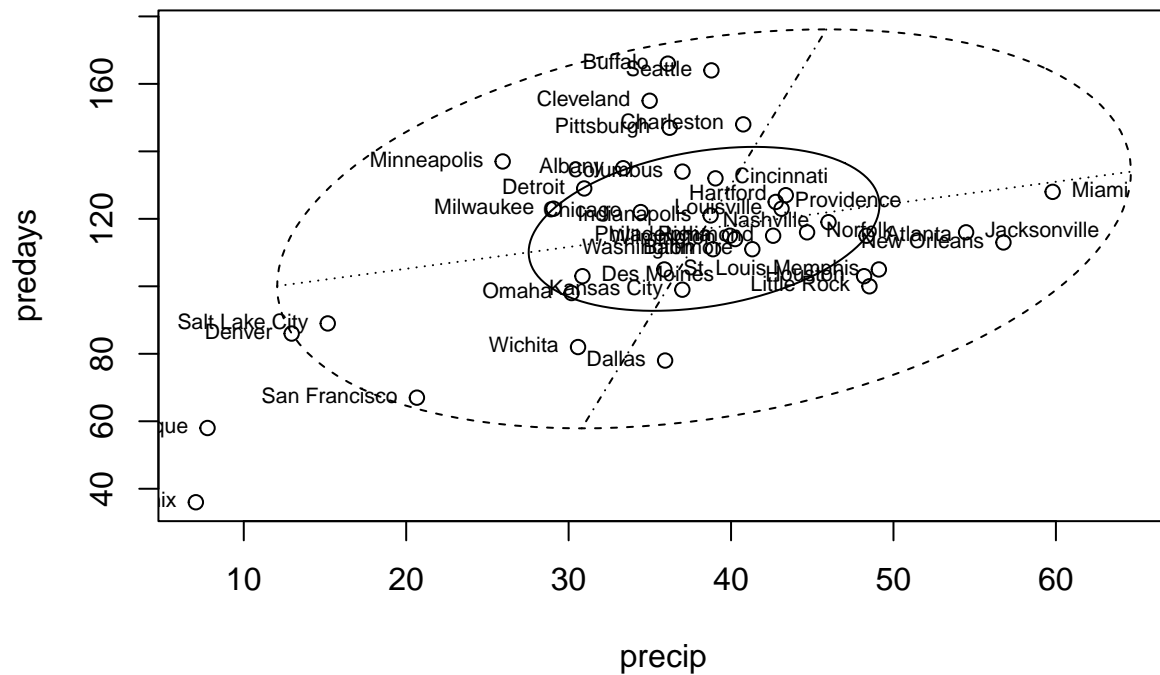












```
cor_set <- c()
for(i in seq(1, 7)){
  for(j in seq(i, 7)) {
    if (i != j) {
      c <- cor(USairpollution[, c(i, j)][,1], USairpollution[, c(i, j)][,2])
      cor_set = append(cor_set, c)
    }
  }
}
cor_set
```

```
## [1] -0.43360020  0.64476873  0.49377958  0.09469045  0.05429434  0.36956363
## [7] -0.19004216 -0.06267813 -0.34973963  0.38625342 -0.43024212  0.95526935
## [13]  0.23794683 -0.03241688  0.13182930  0.21264375 -0.02611873  0.04208319
## [19] -0.01299438  0.16410559  0.49609671
```

2.2

Chi-square plot are usually used when the variables we are plotting are categorical, scatter plots can more directly show the correlation between numeric variables.

```
compare <- function(data, i, j) {
```

```

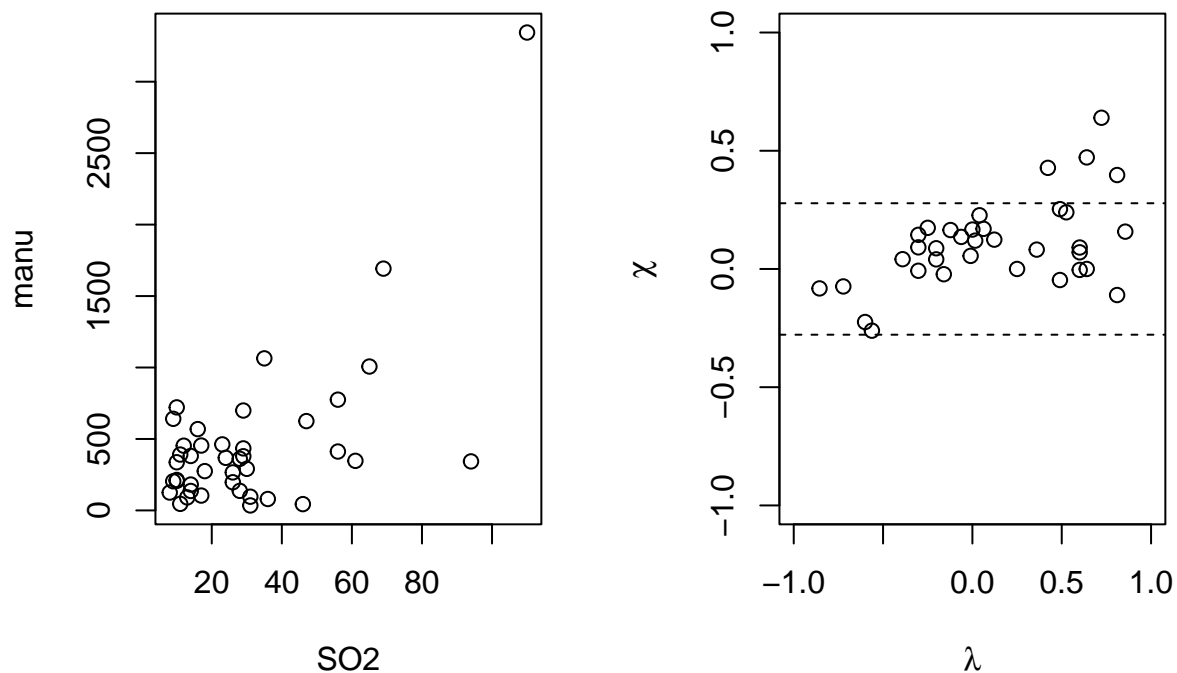
par(mfrow=c(1, 2))

plot(data[[i]], data[[j]], xlab=names(data)[i], ylab=names(data[j]))

chipplot(data[[i]], data[[j]])
}

compare(USairpollution, 1, 3)

```



2.3

I found the graph in 2.17 is more informative since it also shows us how the distribution looks like for each variables.

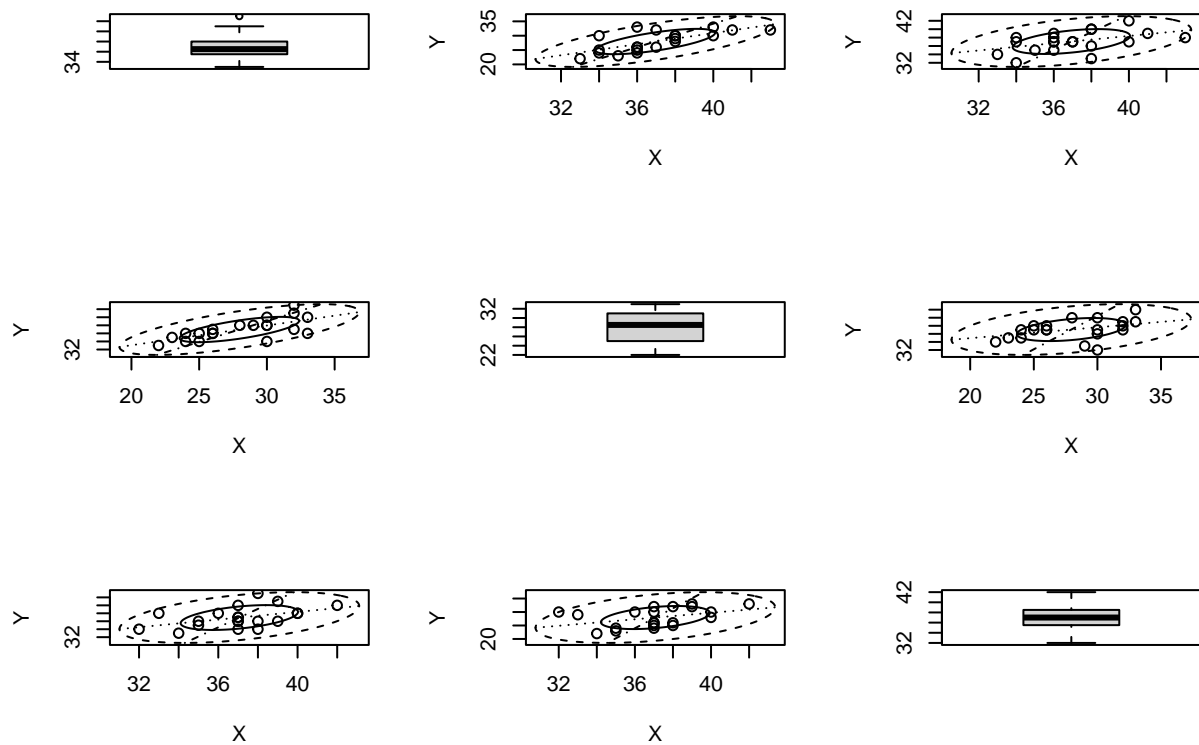
```

measure <- structure(list(V1 = 1:20, V2 = c(34L, 37L, 38L, 36L, 38L, 43L,
      40L, 38L, 40L, 41L, 36L, 36L, 34L, 33L, 36L, 37L, 34L, 36L, 38L,
      35L), V3 = c(30L, 32L, 30L, 33L, 29L, 32L, 33L, 30L, 30L, 32L,
      24L, 25L, 24L, 22L, 26L, 26L, 25L, 26L, 28L, 23L), V4 = c(32L,
      37L, 36L, 39L, 33L, 38L, 42L, 40L, 37L, 39L, 35L, 37L, 37L, 34L,
      38L, 37L, 38L, 37L, 40L, 35L)), .Names = c("V1", "V2", "V3",
      "V4"), class = "data.frame", row.names = c(NA, -20L))
measure <- measure[,-1]
names(measure) <- c("chest", "waist", "hips")

```

```
measure$gender <- gl(2, 10)
levels(measure$gender) <- c("male", "female")
```

```
par(mfrow=c(3, 3))
for (i in seq(1,3)) {
  for (j in seq(1,3)) {
    if(i != j) {
      bvbox(measure[, c(i, j)])
    }
    else {
      boxplot(measure[[i]])
    }
  }
}
```



2.4

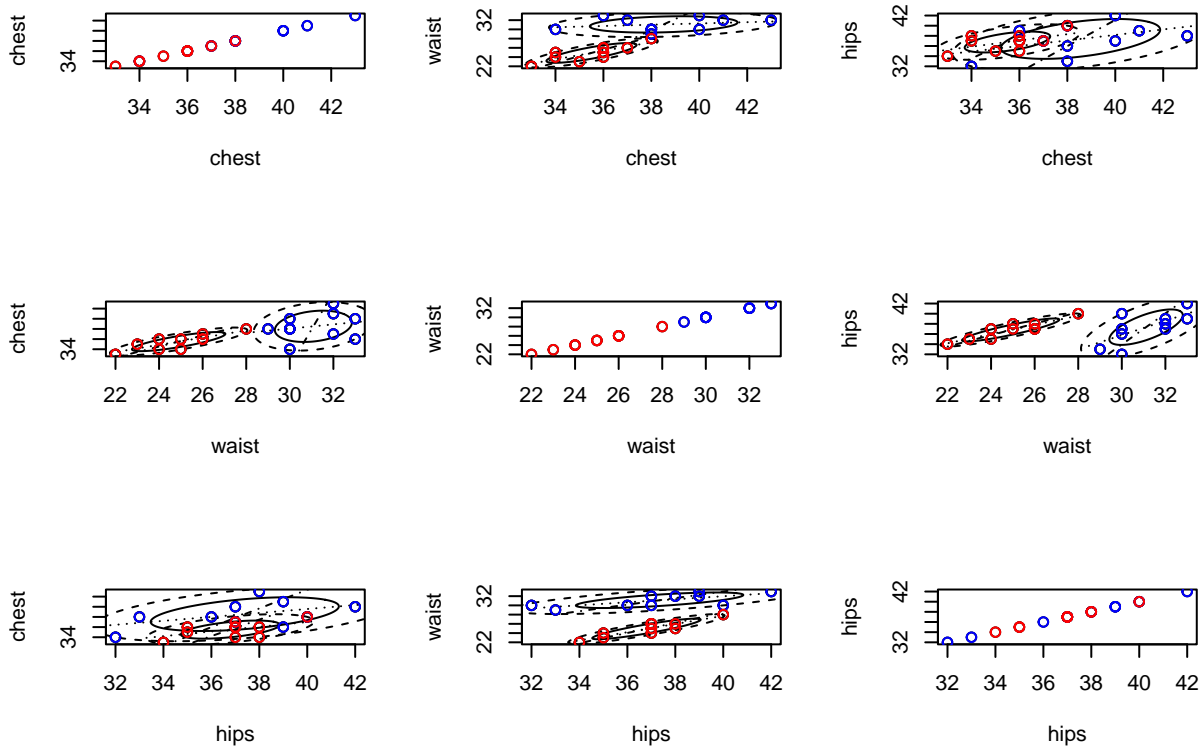
```
par(mfrow=c(3, 3))
for (i in 1:3) {
  for (j in 1:3) {
    plot(measure[, i], measure[, j], xlab = names(measure)[i], ylab=names(measure)[j])
    if(i != j) {
      bvbox(measure[measure$gender == "male", c(i, j)], add=TRUE)
      bvbox(measure[measure$gender == "female", c(i, j)], add=TRUE)
    }
  }
}
```



```

    }
    points(measure[measure$gender == "male", c(i, j)], col="blue")
    points(measure[measure$gender == "female", c(i, j)], col="red")
  }
}

```



2.5

```

library(HSAUR2)
data(pottery, package = "HSAUR2")

```