hw2_5.1.R

guestuser

2020-05-25

```
install.packages("outliers", repos='http://cran.us.r-project.org')
```

```
##
## The downloaded binary packages are in
##  /var/folders/b9/qkd61hd97gz76cvnd_444xp80000gq/T//Rtmpe5Q5aC/downloaded_packages
```

```
library(outliers)
install.packages("ggplot2", repos='http://cran.us.r-project.org')
```

```
##
## The downloaded binary packages are in
##  /var/folders/b9/qkd61hd97gz76cvnd_444xp80000gq/T//Rtmpe5Q5aC/downloaded_packages
```

```
library(ggplot2)


rm(list = ls())
data <- read.table("uscrime.txt",  header = TRUE)

#x = a numeric vector for data values.
#opposite = a logical indicating whether you want to check not the value with largest difference from t
#type= Integer value indicating test variant.
##10 is a test for one outlier (side is detected automatically and can be reversed by opposite paramete
#two.sided= Logical value indicating if there is a need to treat this test as two-sided.
outlier_oneoutlier <- grubbs.test(data[,"Crime"], type = 10, opposite = FALSE, two.sided = FALSE)
outlier_oneoutlier
```

```
##
##  Grubbs test for one outlier
##
## data:  data[, "Crime"]
## G = 2.81287, U = 0.82426, p-value = 0.07887
## alternative hypothesis: highest value 1993 is an outlier
```

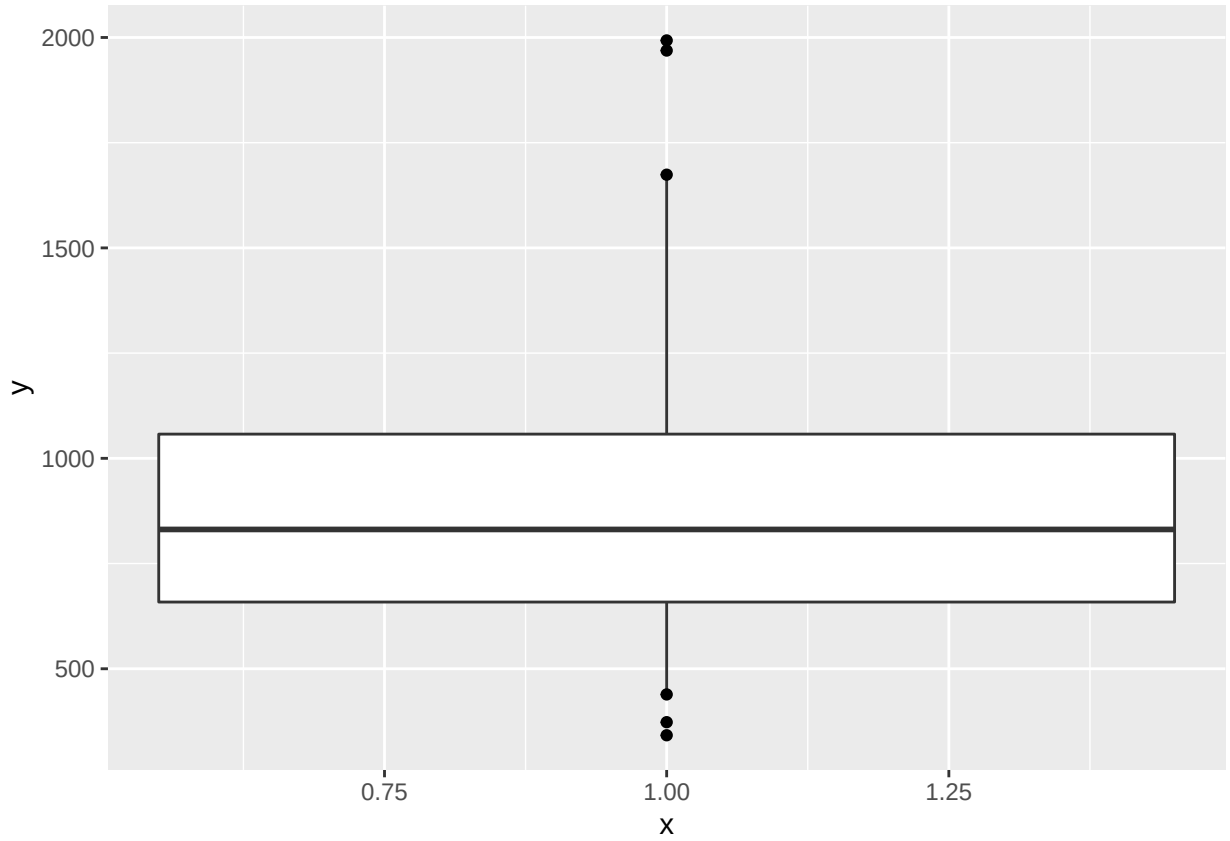
```
cat("Conclustion: The P valve is 0.07887, which means the the highest outlier is 1993.\n")
```

```
## Conclustion: The P valve is 0.07887, which means the the highest outlier is 1993.
```

```
#################################################################################################
#visulization the data to find outliers
# Create data frame with just the crime data
df <- data.frame(x = rep(1, nrow(data)), y = data[,"Crime"])

# Define a function that finds points below and aboce the 5% and 95% quantiles of the data
outliers <- function(x) {
  subset(x, x < quantile(x, 0.05) | quantile(x, 0.95) < x)
}
quant <- function(x) {
  r <- quantile(x, probs = c(0.05, 0.25, 0.5, 0.75, 0.95))
  names(r) <- c("ymin", "lower", "middle", "upper", "ymax")
  r
}

# Create the box-and-whisker plot
ggplot(df, aes(x, y)) +
  stat_summary(fun.data = quant, geom="boxplot") +
  stat_summary(fun = outliers, geom="point")
```




```
cat("Boxplot shows there are 2 high outliers and 2 low outliers")
```

```
## Boxplot shows there are 2 high outliers and 2 low outliers
```