R Homework 6

Fisher Ankney

Statistics 5193

Question 1

Consider the social media data on Brightspace. The objective of this question is to determine if introverted students tend to send or receive more texts. To accomplish this objective, you will construct a plot of TxtRec – TxtSent on the y axis and Introvert on the x axis. For full credit the plot must satisfy each of the following. At least one separate line of code should be provided for each letter.

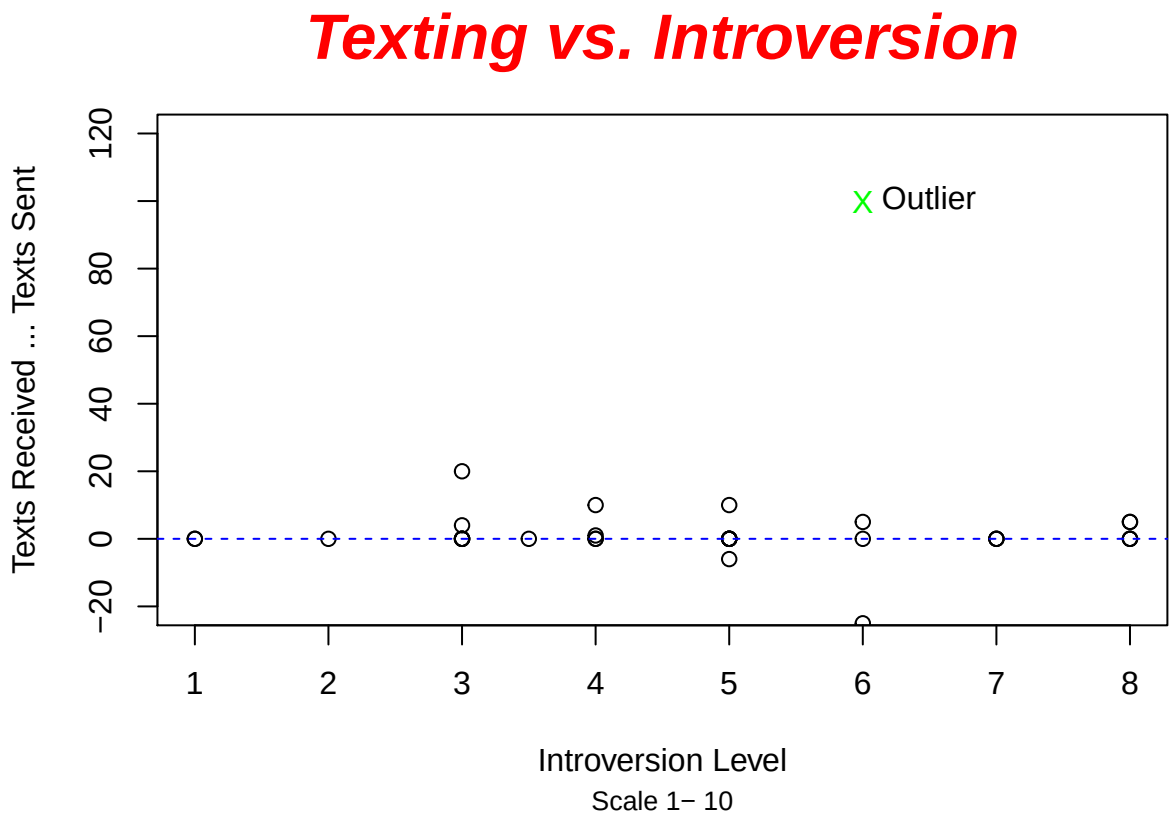
- Generate the plot with the y axis labeled "Texts Received – Texts Sent" and the x axis labeled "Introversion Level"
- Create a subtitle that is 80% of the default size that reads "Scale: 1 – 10"
- Create a title reading "Texting vs. Introversion" that is bold, in italics, and red. It should also be twice the default size.
- Add a horizontal blue dashed line at 0 on the plot.
- Write "Outlier" to the right of the outlier and plot the outlier with a different plotting character and color. For full credit, you must recreate the plot without the outlier and then add the point to the plot. That is, don't write over the top of the circle.

```
library(readxl)
StudentData <- read_excel("/Users/fisher/Documents/data_science/r_stat_5193/data/StudentData.xlsx")
```

```
StudentDataNoOut <- StudentData[-28,]
StudentDataOut <- StudentData[28,]
```

```
plot(StudentDataNoOut$TxtRec
    - StudentDataNoOut$TxtSent
    ~ StudentDataNoOut$Introvert,
    ylab = "Texts Received - Texts Sent",
    xlab = "Introversion Level",
    ylim = c(-20, 120))

title(main = "Texting vs. Introversion",
      cex.main = 2,
      font.main = 4,
      col.main = 'red')

title(sub = "Scale 1- 10",
      cex.sub = .8)

lines(abline(h = 0,
             col = 'blue',
             lty = 2))

points(StudentDataOut$TxtRec
       - StudentDataOut$TxtSent
       ~ StudentDataOut$Introvert,
       pch = "X", col = 'green')
```

```
text(StudentDataOut$TxtRec
     - StudentDataOut$TxtSent
     ~ StudentDataOut$Introvert,
     label = 'Outlier',
     pos=4)
```



Question 2

Now we'll recreate the plot to illustrate the effect of the outlier.

- Recreate the plot above but with the outlier removed. The title should now also read "Outlier Analysis" and the y axis should be from -25 to 20.
- Add a thick red solid least squares regression line that uses all the data.
- Add a thick black dashed least squares regression line that is fit to the data after the outlier is removed.
- Add a legend in the bottom left corner of the plot that has title "Fits". A red solid line should be drawn and labeled "With Outlier" and a black dashed line should be labeled "Without Outlier". The legend should be inset a little.

```
plot(StudentDataNoOut$TxtRec
    - StudentDataNoOut$TxtSent
    ~ StudentDataNoOut$Introvert,
    ylab = "Texts Received - Texts Sent",
    xlab = "Introversion Level",
    ylim = c(-25, 20))
```

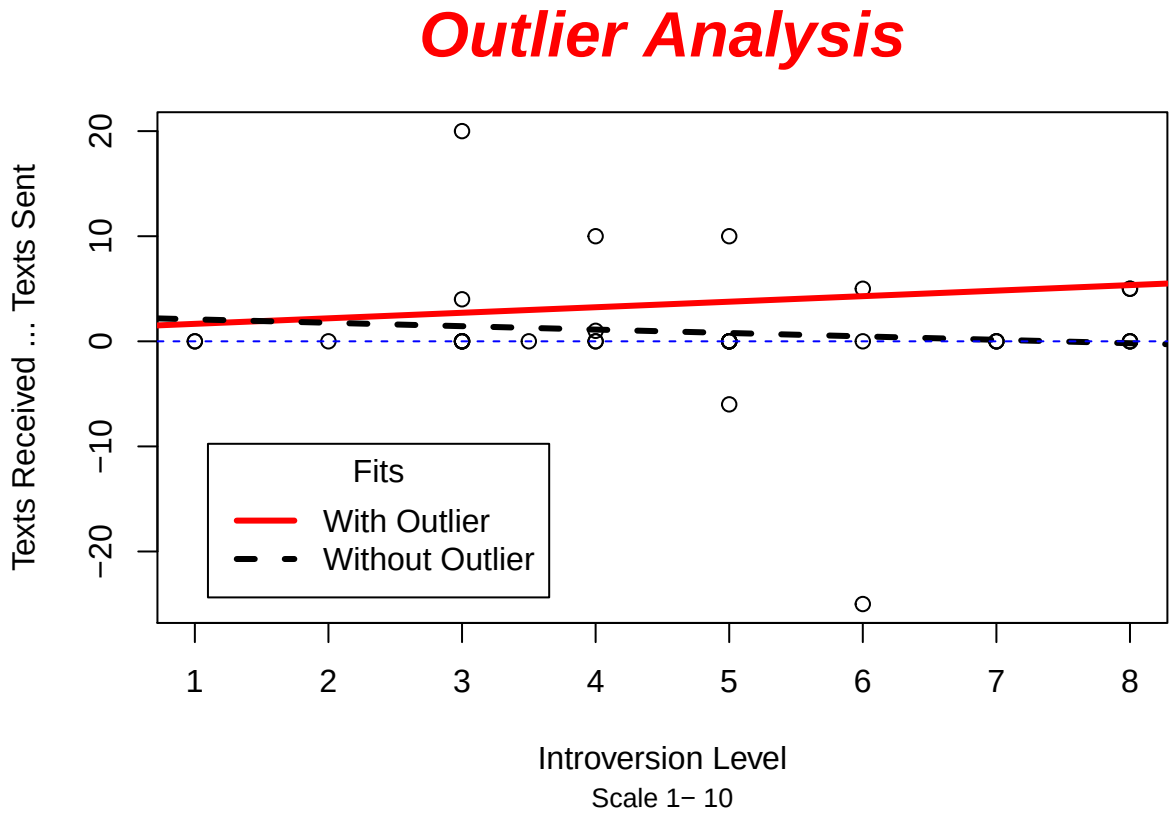
```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <93>
```

```
title(main = "Outlier Analysis",
      cex.main = 2,
      font.main = 4,
      col.main = 'red')

title(sub = "Scale 1- 10",
      cex.sub = .8)

no_outlier_reg <- lm(StudentDataNoOut$TxtRec
                     - StudentDataNoOut$TxtSent
                     ~ StudentDataNoOut$Introvert)

outlier_reg <- lm(StudentData$TxtRec
                     - StudentData$TxtSent
                     ~ StudentData$Introvert)

lines(abline(outlier_reg,
             col = 'red',
             lwd=3))

lines(abline(no_outlier_reg,
             col = 'black',
             lwd=3,
             lty=2))
```

```
lines(abline(h = 0,
             col = 'blue',
             lty = 2))

legend("bottomleft",
       inset = 0.05,
       title = 'Fits',
       c("With Outlier","Without Outlier"),
       lty=c(1,2),
       lwd=c(3,3),
       col=c("red", "black"))
```



Question 3

We will create a plot of the same data with the same x and y labels and title as in part 1, but also identify Pinterest users from non-Pinterest users.

- Generate the plot with the y axis labeled "Texts Received – Texts Sent" and the x axis labeled "Introversion Level" but use red triangles to plot Snapchat users and black circles to plot students without Snapchat. Hint: try running as.numeric(Snapchat)
- Add the title "Texting vs. Introversion vs. Snapchat".
- Add a legend in the top left corner of the plot with title "Snapchat" and labels "Y" and "N" with the appropriate plotting characters and colors. The legend should be inset a little.

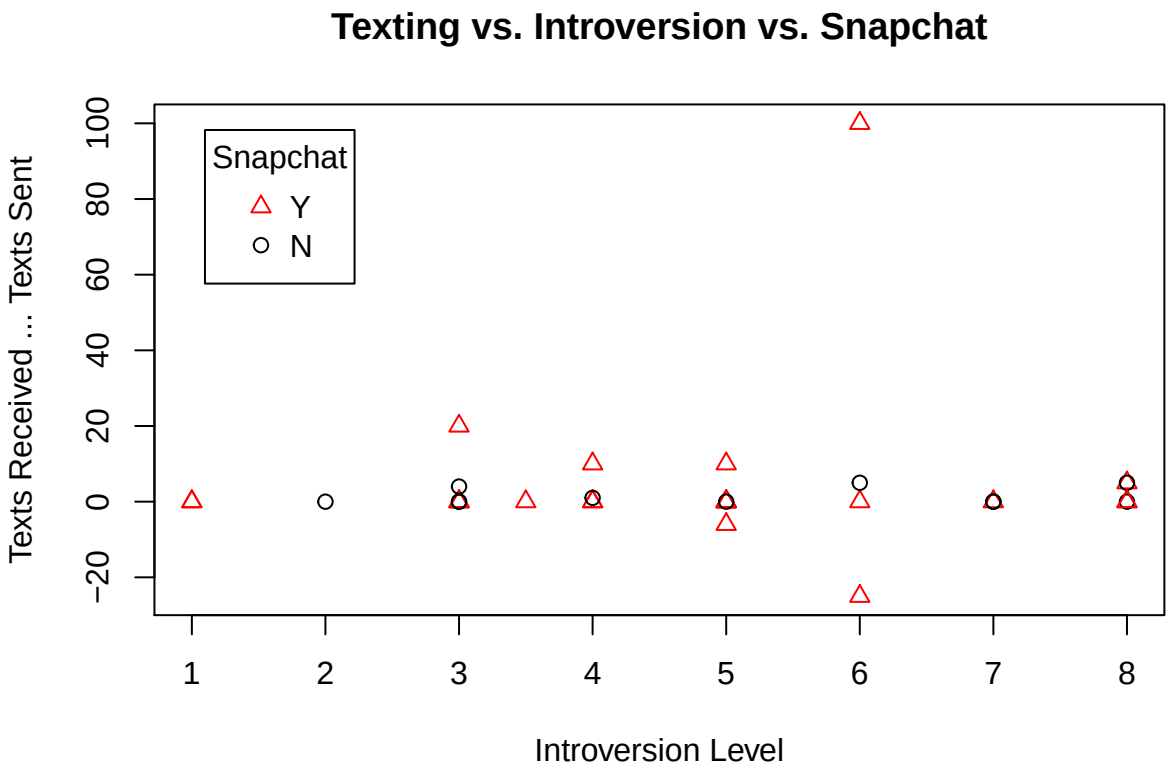
```
plot(StudentData$TxtRec
    - StudentData$TxtSent
    ~ StudentData$Introvert,
    pch = c(1,2)[as.numeric(as.factor(StudentData$Snapchat))],
    col = c('black','red')[as.numeric(as.factor(StudentData$Snapchat))],
    ylab = "Texts Received - Texts Sent",
    xlab = "Introversion Level",
    )
```

```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in title(...): conversion failure on 'Texts Received - Texts Sent'
## in 'mbcsToSbcs': dot substituted for <93>
```

```
title(main = "Texting vs. Introversion vs. Snapchat")

legend("topleft",
       inset = 0.05,
       title='Snapchat',
       c("Y","N"),
       pch = c(2,1),
       col = c('red', 'black'))
```


Texting vs. Introversion vs. Snapchat
Texts Received ... Texts Sent
Introversion Level
Snapchat
Y
N
−20
0
20
40
60
80
100
1
2
3
4
5
6
7
8