HW3_tdolkar

Due Wednesday Sep 30

Tsering Dolkar

9/24/2020

Problem 1

I did primer on Rstudio cloud.

Problem 2

Created the Rmd file

Problem 3

Problem 4

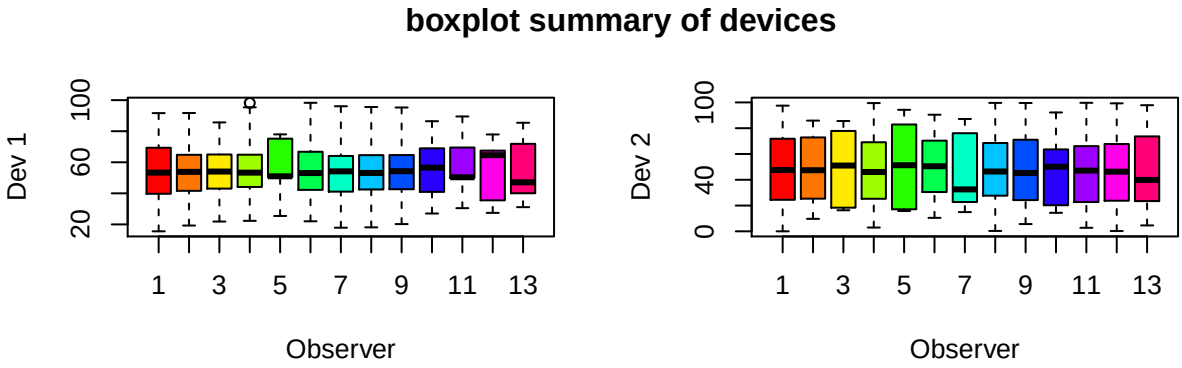
Problem 5

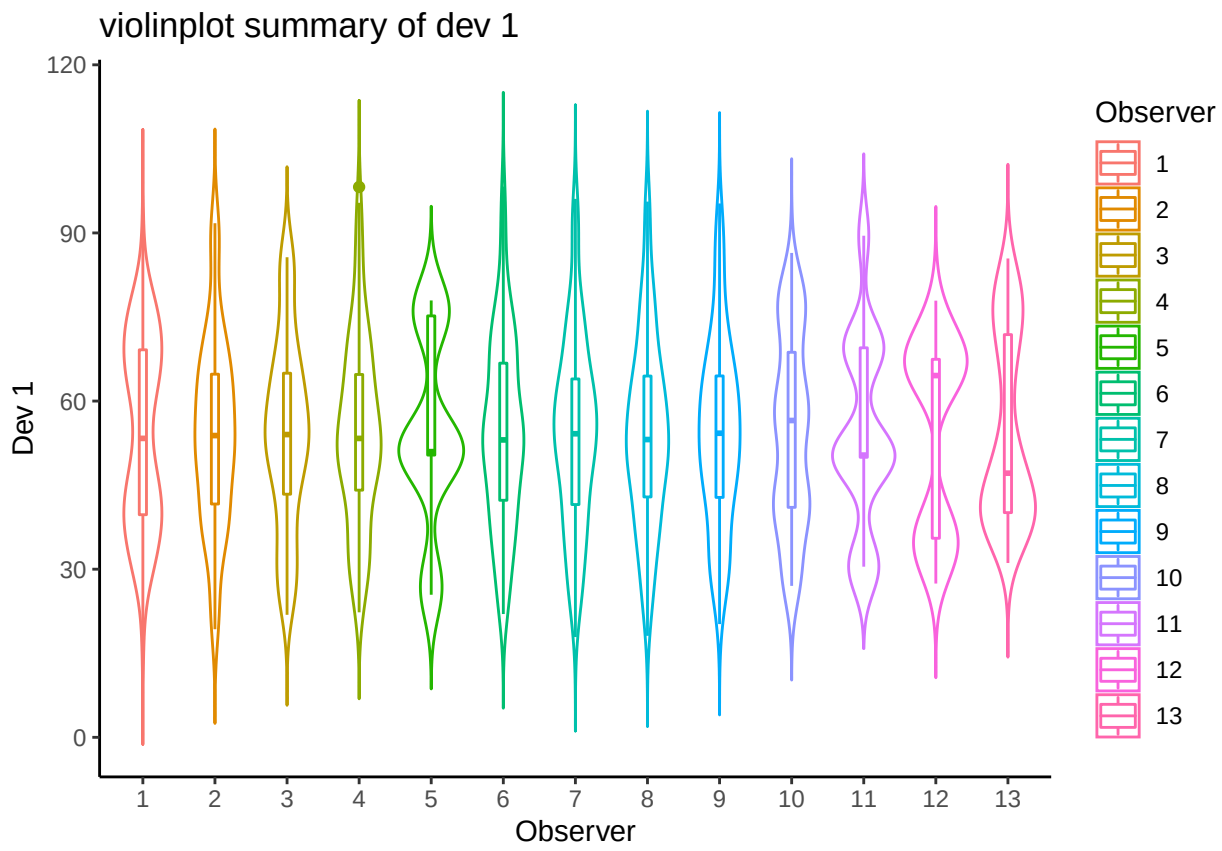
#dplyr + summarize + group_by

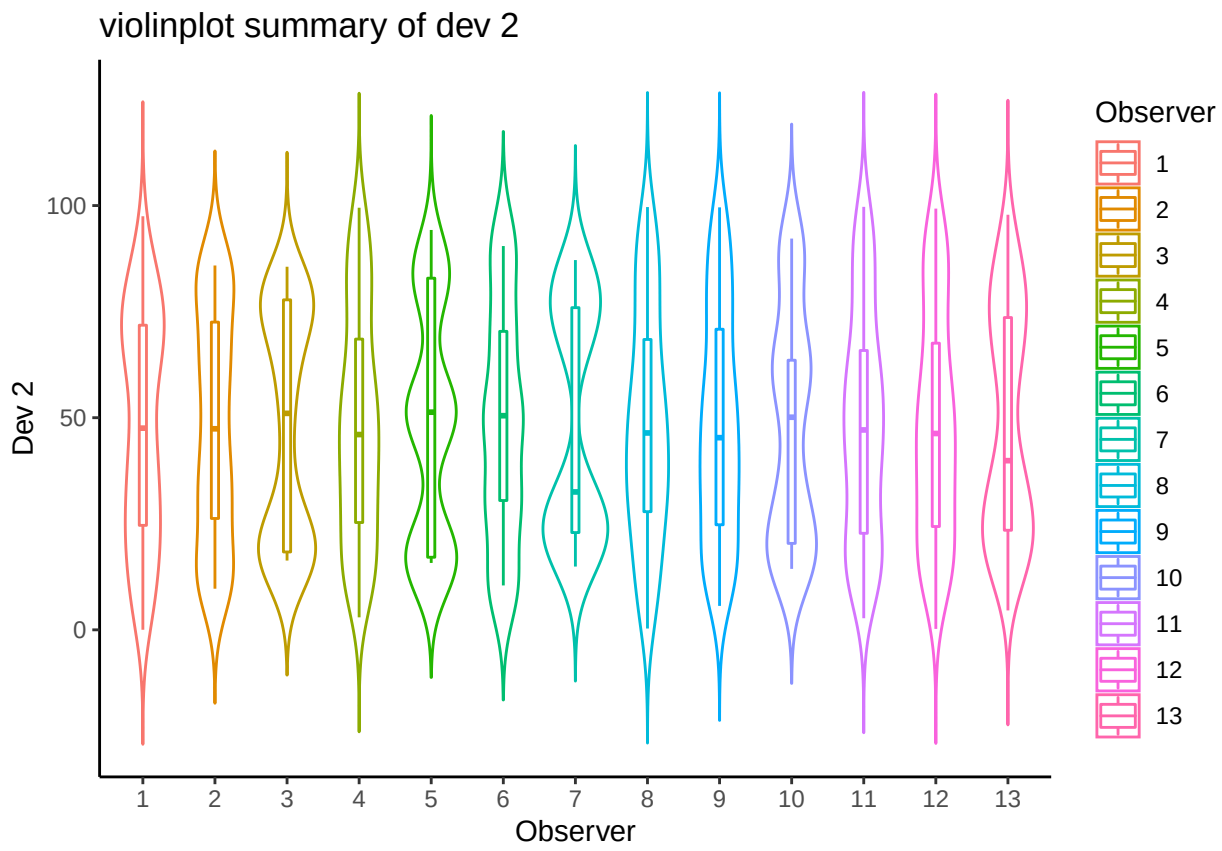
```
summarise_data <- function(our_data){
  #function to calculate summary of a dataframe, returns a vector of the summary
  observer_summary <- double(length = 5)
  for(i in 1:5){
    if(i == 1)
      observer_summary[i] <- mean(our_data[,1])
    if(i == 2)
      observer_summary[i] <- mean(our_data[,2])
    if(i == 3)
      observer_summary[i] <- sd(our_data[,1])
    if(i == 4)
      observer_summary[i] <- sd(our_data[,2])
    if(i == 5)
      observer_summary[i] <- cor(our_data[,1], our_data[,2])
  }
  return(observer_summary)
}
```

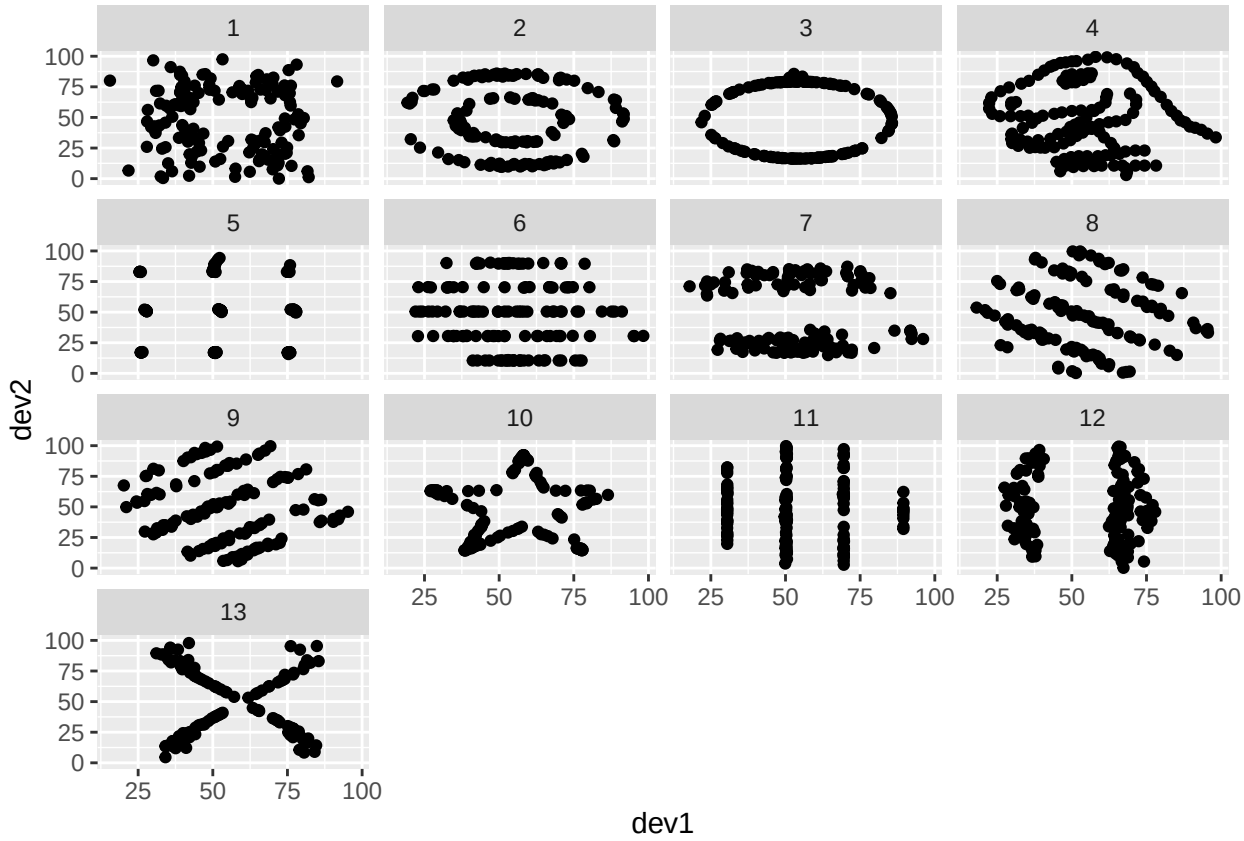
```r
# We'll find the summary for each observer:
# We will be returned a vector of length 5 for each of 13 observers from the function
# summarise_data such that
# 1. mean of dev 1
# 2. mean of dev 2
# 3. standard dev of dev 1
# 4. standard dev of dev 2
# 5. correlation between dev 1 and 2
# are the values respectively in a row.

#separate the data into vectors to make it less confusing to work with first
observer <- observations$Observer
dev1 <- observations$dev1
dev2 <- observations$dev2

#initialize vectors we will need:
dev1_by_observer <- double(0)
dev2_by_observer <- double(0)
summary_statistics <- data.frame()
colnames(summary_statistics) <-

#we have a nested for loop here. The outer for loop keeps track of observer 1 to 13 and
#the inside for loop looks for all the data by the said observer in the dataset from top to bottom
#once.
for(i in 1:13){
   track_observer <- i
   for(j in 1:length(observer)){
     if(track_observer == observer[j]){
     dev1_by_observer <- c(dev1_by_observer, dev1[j])
     dev2_by_observer <- c(dev2_by_observer, dev2[j])
     }
   }
   raw_data <- data.frame(dev1_by_observer, dev2_by_observer)
   colnames(raw_data) <- c("dev1", "dev2")
   summary_statistics_each_observer <- cbind(rep(i, 5), summarise_data(raw_data))
   summary_statistics <- rbind(summary_statistics, summary_statistics_each_observer)
}
summary_statistics <- data.frame(rep(c("Mean_dev1", "Mean_dev2", "SD_dev1", "SD_dev2",
                                       "Cor_dev1_dev2"), 13), summary_statistics)
colnames(summary_statistics) <- c('V1','Observer','V3')
new_summary_statistics <- summary_statistics %>%
                          spread(key = V1, value = V3)
# reorder by column name
new_summary_statistics <- new_summary_statistics[c("Observer", "Mean_dev1", "Mean_dev2",
                                                   "SD_dev1", "SD_dev2", "Cor_dev1_dev2")]
new_summary_statistics <- kable(new_summary_statistics)
```

```
## [1] "boxplot summary of devices"
```


boxplot summary of devices
Dev 1
20
60
100
1
3
5
7
9
11
13
Observer
Dev 2
0
40
100
1
3
5
7
9
11
13
Observer


violinplot summary of dev 1
Dev 1
120
90
60
30
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Observer
Observer
1
2
3
4
5
6
7
8
9
10
11
12
13


violinplot summary of dev 2
Dev 2
100
50
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Observer
Observer
1
2
3
4
5
6
7
8
9
10
11
12
13


1
2
3
4
5
6
7
8
9
10
11
12
13
dev2
dev1
0
25
50
75
100