Revisitng R (Extroverts and productivity)

Belle Joy Stevens

2022-05-18

## Extrovert Rating

The extrovert status makes a little difference in revenue and overall productivity

library(readr)  
project <- read\_csv("C:/Users/belle/OneDrive/Desktop/Analytics/project.txt")  
library(tidyverse)  
glimpse(project)

## Rows: 1,512  
## Columns: 8  
## $ education <chr> "HS", "HS", "HS", "COL", "HS", "HS", "HS", "HS", "COL", "C…  
## $ absences <dbl> 6, 3, 2, 1, 4, 1, 4, 3, 6, 6, 2, 6, 5, 1, 4, 3, 1, 3, 5, 4…  
## $ extrovert <chr> "MED", "HIGH", "MED", "LOW", "LOW", "MED", "HIGH", "HIGH",…  
## $ prior\_sales <dbl> 3, 0, 8, 2, 4, 0, 0, 0, 5, 2, 3, 5, 5, 2, 4, 6, 5, 5, 4, 0…  
## $ recruit <chr> "FAIR", "FAIR", "WEB", "WEB", "REF", "REF", "WEB", "WEB", …  
## $ grammar <dbl> 2, 3, 8, 3, 3, 6, 3, 3, 5, 4, 2, 8, 7, 7, 4, 5, 8, 6, 6, 6…  
## $ first <dbl> 223, 191, 307, 155, 258, 261, 221, 328, 213, 300, 237, 280…  
## $ revenue <dbl> 186, 417, 479, 250, 200, 372, 387, 428, 313, 535, 292, 340…

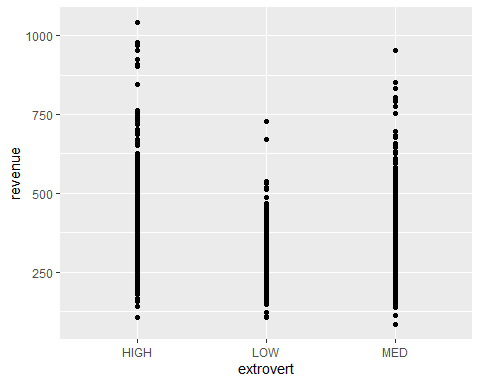
library(reader)  
install.packages("reader")

## Warning: package 'reader' is in use and will not be installed

library(reader)  
glimpse(project)

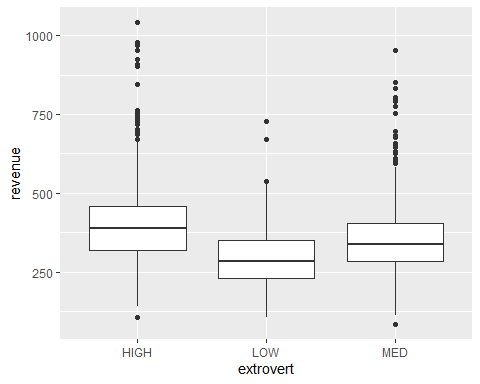
## Rows: 1,512  
## Columns: 8  
## $ education <chr> "HS", "HS", "HS", "COL", "HS", "HS", "HS", "HS", "COL", "C…  
## $ absences <dbl> 6, 3, 2, 1, 4, 1, 4, 3, 6, 6, 2, 6, 5, 1, 4, 3, 1, 3, 5, 4…  
## $ extrovert <chr> "MED", "HIGH", "MED", "LOW", "LOW", "MED", "HIGH", "HIGH",…  
## $ prior\_sales <dbl> 3, 0, 8, 2, 4, 0, 0, 0, 5, 2, 3, 5, 5, 2, 4, 6, 5, 5, 4, 0…  
## $ recruit <chr> "FAIR", "FAIR", "WEB", "WEB", "REF", "REF", "WEB", "WEB", …  
## $ grammar <dbl> 2, 3, 8, 3, 3, 6, 3, 3, 5, 4, 2, 8, 7, 7, 4, 5, 8, 6, 6, 6…  
## $ first <dbl> 223, 191, 307, 155, 258, 261, 221, 328, 213, 300, 237, 280…  
## $ revenue <dbl> 186, 417, 479, 250, 200, 372, 387, 428, 313, 535, 292, 340…

qplot(extrovert, revenue, data = project)



Generic qqplots are not very detailed. Lets try boxplots

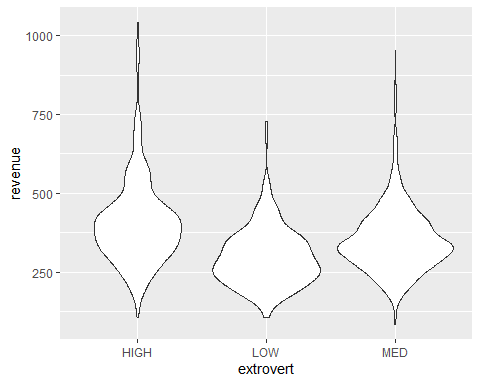
qplot(extrovert, revenue, data = project, geom = 'boxplot')



This gives you an idea about the median and being an extrovert is’nt doing much because the distribution above and below the median looks almost the same

Lets replace boxplots with violin graph

qplot(extrovert, revenue, data = project, geom = 'violin')



This is a little more detailed, the wide area is the median.

Using other useful functions of R like mean, median and group\_by (their education levels), filter (in this case filtering employees with no prior experience)

project %>% summarize(avg = mean(revenue))

## # A tibble: 1 × 1  
## avg  
## <dbl>  
## 1 360.

project %>% summarize(avg = mean(revenue), med = median(revenue), count = n())

## # A tibble: 1 × 3  
## avg med count  
## <dbl> <dbl> <int>  
## 1 360. 342 1512

project %>% group\_by(education) %>% summarize(avg = mean(revenue), med = median(revenue), count = n())

## # A tibble: 3 × 4  
## education avg med count  
## <chr> <dbl> <dbl> <int>  
## 1 COL 374. 360. 226  
## 2 GED 373. 352 260  
## 3 HS 353. 338 1026

no\_prior = project %>% filter(prior\_sales == 0)

Next, we find the quantiles and bottom and top employees and find a summary of it

q10 = quantile(project$revenue, 0.1)  
q10

## 10%   
## 227

q90 = quantile(project$revenue, 0.9)  
q90

## 90%   
## 496.9

top10 = project %>% filter(revenue >= q90)  
top10 = project %>% filter(revenue >= q90)  
bottom10 = project %>% filter(revenue <= q10)  
top10 = project %>% filter(revenue >= q90)  
bottom10

## # A tibble: 155 × 8  
## education absences extrovert prior\_sales recruit grammar first revenue  
## <chr> <dbl> <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 HS 6 MED 3 FAIR 2 223 186  
## 2 HS 4 LOW 4 REF 3 258 200  
## 3 HS 5 HIGH 6 WEB 9 112 207  
## 4 HS 6 HIGH 4 WEB 5 112 142  
## 5 HS 3 MED 2 WEB 5 150 181  
## 6 HS 3 MED 0 FAIR 3 347 210  
## 7 HS 11 HIGH 2 WEB 3 214 219  
## 8 HS 4 LOW 7 WEB 6 109 162  
## 9 HS 6 MED 0 REF 3 186 190  
## 10 HS 2 MED 8 WEB 3 130 193  
## # … with 145 more rows

summary(top10)

## education absences extrovert prior\_sales   
## Length:152 Min. : 0.000 Length:152 Min. : 0.000   
## Class :character 1st Qu.: 1.000 Class :character 1st Qu.: 5.000   
## Mode :character Median : 2.000 Mode :character Median : 6.000   
## Mean : 2.697 Mean : 5.342   
## 3rd Qu.: 4.000 3rd Qu.: 7.000   
## Max. :11.000 Max. :15.000   
## recruit grammar first revenue   
## Length:152 Min. :0.000 Min. :115.0 Min. : 497.0   
## Class :character 1st Qu.:4.000 1st Qu.:180.5 1st Qu.: 533.8   
## Mode :character Median :6.000 Median :217.5 Median : 573.5   
## Mean :5.789 Mean :231.2 Mean : 613.3   
## 3rd Qu.:7.000 3rd Qu.:276.0 3rd Qu.: 659.0   
## Max. :9.000 Max. :493.0 Max. :1042.0

summary(bottom10)

## education absences extrovert prior\_sales   
## Length:155 Min. : 0.000 Length:155 Min. : 0.000   
## Class :character 1st Qu.: 3.000 Class :character 1st Qu.: 0.000   
## Mode :character Median : 4.000 Mode :character Median : 4.000   
## Mean : 4.077 Mean : 3.303   
## 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. :11.000 Max. :15.000   
## recruit grammar first revenue   
## Length:155 Min. :0.000 Min. :109.0 Min. : 84.0   
## Class :character 1st Qu.:2.000 1st Qu.:157.5 1st Qu.:181.0   
## Mode :character Median :4.000 Median :193.0 Median :203.0   
## Mean :4.355 Mean :206.1 Mean :195.3   
## 3rd Qu.:7.000 3rd Qu.:245.0 3rd Qu.:218.0   
## Max. :9.000 Max. :393.0 Max. :227.0

we see that extrovert variable is recognized as character and not a string and that doesnt add any value to the data. Therefore lets add a new row where LOW, MED and HIGH will be strings and lets see how that changes the data

ext\_levels = c('LOW', 'MED', 'HIGH')  
project = project %>% mutate(ext\_factors = factor(extrovert, levels = ext\_levels))  
glimpse(project)

## Rows: 1,512  
## Columns: 9  
## $ education <chr> "HS", "HS", "HS", "COL", "HS", "HS", "HS", "HS", "COL", "C…  
## $ absences <dbl> 6, 3, 2, 1, 4, 1, 4, 3, 6, 6, 2, 6, 5, 1, 4, 3, 1, 3, 5, 4…  
## $ extrovert <chr> "MED", "HIGH", "MED", "LOW", "LOW", "MED", "HIGH", "HIGH",…  
## $ prior\_sales <dbl> 3, 0, 8, 2, 4, 0, 0, 0, 5, 2, 3, 5, 5, 2, 4, 6, 5, 5, 4, 0…  
## $ recruit <chr> "FAIR", "FAIR", "WEB", "WEB", "REF", "REF", "WEB", "WEB", …  
## $ grammar <dbl> 2, 3, 8, 3, 3, 6, 3, 3, 5, 4, 2, 8, 7, 7, 4, 5, 8, 6, 6, 6…  
## $ first <dbl> 223, 191, 307, 155, 258, 261, 221, 328, 213, 300, 237, 280…  
## $ revenue <dbl> 186, 417, 479, 250, 200, 372, 387, 428, 313, 535, 292, 340…  
## $ ext\_factors <fct> MED, HIGH, MED, LOW, LOW, MED, HIGH, HIGH, MED, HIGH, HIGH…

bottom10 = project %>% filter(revenue <= q10)  
top10 = project %>% filter(revenue >= q90)  
summary(top10)

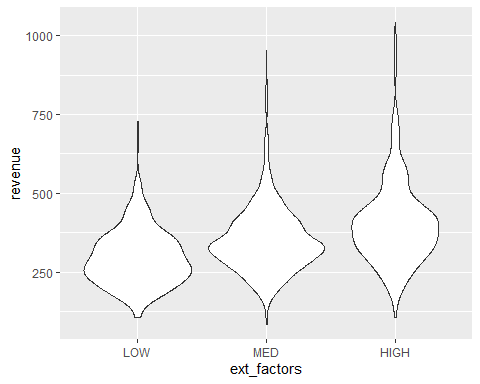
## education absences extrovert prior\_sales   
## Length:152 Min. : 0.000 Length:152 Min. : 0.000   
## Class :character 1st Qu.: 1.000 Class :character 1st Qu.: 5.000   
## Mode :character Median : 2.000 Mode :character Median : 6.000   
## Mean : 2.697 Mean : 5.342   
## 3rd Qu.: 4.000 3rd Qu.: 7.000   
## Max. :11.000 Max. :15.000   
## recruit grammar first revenue   
## Length:152 Min. :0.000 Min. :115.0 Min. : 497.0   
## Class :character 1st Qu.:4.000 1st Qu.:180.5 1st Qu.: 533.8   
## Mode :character Median :6.000 Median :217.5 Median : 573.5   
## Mean :5.789 Mean :231.2 Mean : 613.3   
## 3rd Qu.:7.000 3rd Qu.:276.0 3rd Qu.: 659.0   
## Max. :9.000 Max. :493.0 Max. :1042.0   
## ext\_factors  
## LOW : 7   
## MED :58   
## HIGH:87   
##   
##   
##

summary(bottom10)

## education absences extrovert prior\_sales   
## Length:155 Min. : 0.000 Length:155 Min. : 0.000   
## Class :character 1st Qu.: 3.000 Class :character 1st Qu.: 0.000   
## Mode :character Median : 4.000 Mode :character Median : 4.000   
## Mean : 4.077 Mean : 3.303   
## 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. :11.000 Max. :15.000   
## recruit grammar first revenue ext\_factors  
## Length:155 Min. :0.000 Min. :109.0 Min. : 84.0 LOW :60   
## Class :character 1st Qu.:2.000 1st Qu.:157.5 1st Qu.:181.0 MED :68   
## Mode :character Median :4.000 Median :193.0 Median :203.0 HIGH:27   
## Mean :4.355 Mean :206.1 Mean :195.3   
## 3rd Qu.:7.000 3rd Qu.:245.0 3rd Qu.:218.0   
## Max. :9.000 Max. :393.0 Max. :227.0

No we can see how well the extroverts are doing when the variable became string. Now lets redo the violin graph. Instead of extrovert we replace with ext\_factors on the X-axis

qplot(ext\_factors, revenue, data = project, geom = 'violin')



This makes the data much more meaningful