Creating Composite Scores in R

Andy Grogan-Kaylor

2020-11-16

Contents

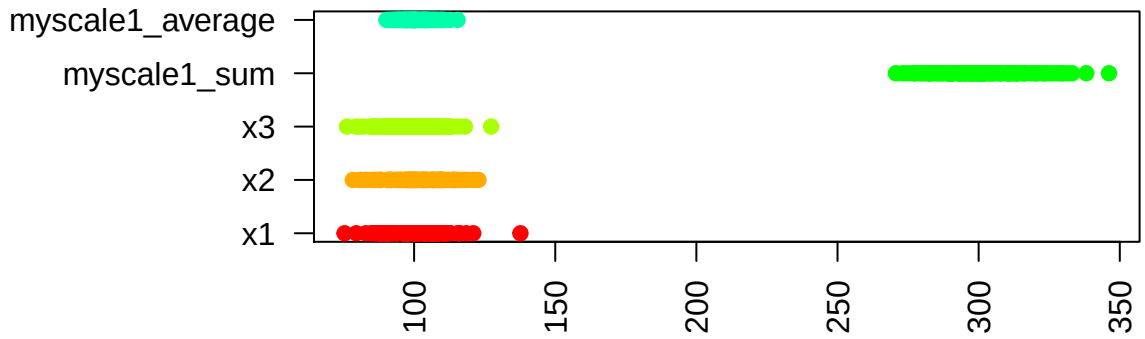
All Variables on Same Scale

Simulate Some Data

```
N <- 100 # sample size

x1 <- rnorm(N, 100, 10)

x2 <- rnorm(N, 100, 10)

x3 <- rnorm(N, 100, 10)
```

Make A Scale

```
myscale1_sum <- x1 + x2 +x3 # scale is sum of variables

myscale1_average <- (x1 + x2 + x3)/3 # scale is average of variables

mydata <- data.frame(x1, x2, x3, myscale1_sum, myscale1_average)

par(mar = c(3, 10, 1, 1)) # graph margins

stripchart(mydata, col = rainbow(9), pch=19, las = 2) # stripchart
```



When variables are on the same scale, their average is on that same scale, and all of the variables *equally* contribute to the sum.

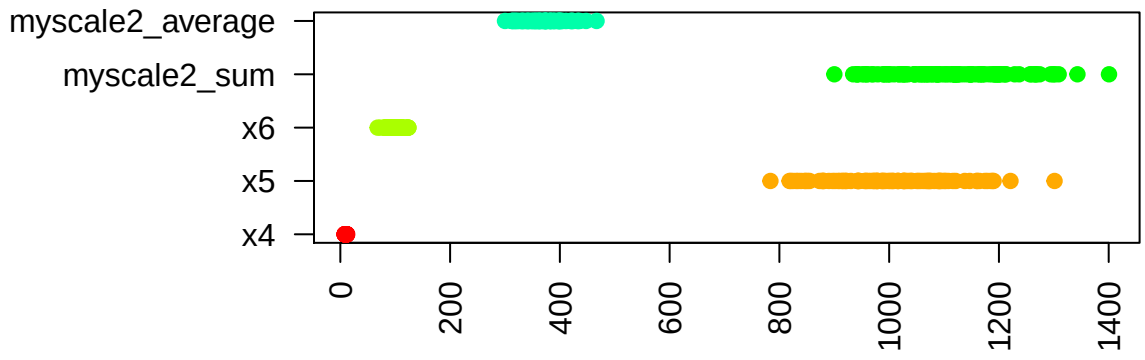
All Variables on Different Scales

Simulate Some Data

```
N <- 100 # sample size

x4 <- rnorm(N, 10, 1)

x5 <- rnorm(N, 1000, 100)

x6 <- rnorm(N, 100, 10)
```

Make A Scale

```
myscale2_sum <- x4 + x5 +x6 # scale is sum of variables

myscale2_average <- (x4 + x5 + x6)/3 # scale is average of variables

mydata2 <- data.frame(x4, x5, x6, myscale2_sum, myscale2_average)

par(mar = c(3, 10, 1, 1)) # graph margins

stripchart(mydata2, col = rainbow(9), pch=19, las = 2) # stripchart
```



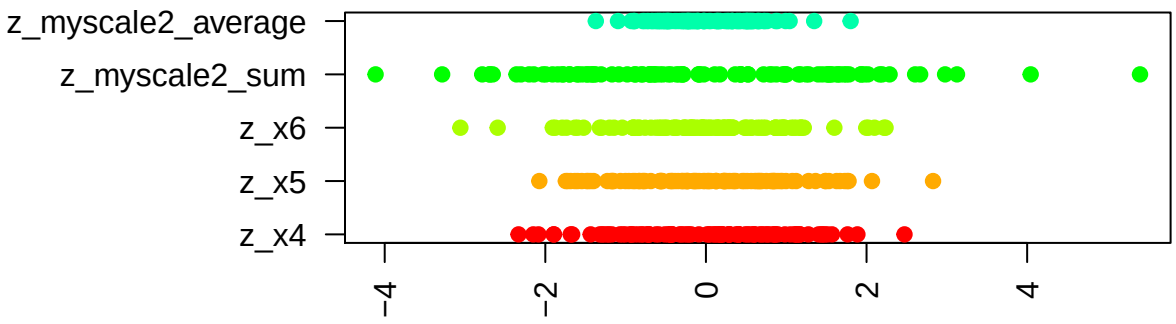
When variables are on different scales, their average is often not particularly reflective of any single variable, and all of the variables *differentially* contribute to the sum.

Solution: z-score

$$x_z = \frac{x - \bar{x}}{sd_x}$$

In reproducing the example below, you may wish to use the `dataset$variable` notation.

```
z_x4 <- (x4 - mean(x4))/ sd(x4) # standardize

z_x5 <- (x5 - mean(x5))/ sd(x5) # standardize

z_x6 <- (x6 - mean(x6))/ sd(x6) # standardize

z_myscale2_sum <- z_x4 + z_x5 + z_x6 # scale is sum of variables

z_myscale2_average <- (z_x4 + z_x5 + z_x6)/3 # scale is average of variables

mydata3 <- data.frame(z_x4, z_x5, z_x6, z_myscale2_sum, z_myscale2_average)

par(mar = c(3, 10, 1, 1)) # graph margins

stripchart(mydata3, col = rainbow(9), pch=19, las = 2) # stripchart
```



> When variables are z-scored, the average is reflective of all the variables, and all variables contribute equallly to the sum.