R Project 1 - Hello R

Aziz Surani

1/23/2020

### Load packages

library(tidyverse)   
library(datasauRus)

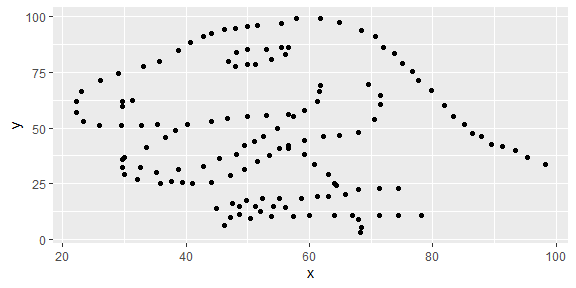
### Exercise 1

the dataset has 3 columns. A data frame with 1846 rows and 3 variables namely dataset, X: x-values, Y: y-values.

### Exercise 2

First let’s plot the data in the dino dataset:

dino\_data <- datasaurus\_dozen %>%  
 filter(dataset == "dino")  
ggplot(data = dino\_data, mapping = aes(x = x, y = y)) +  
 geom\_point()



And next calculate the correlation between x and y in this dataset:

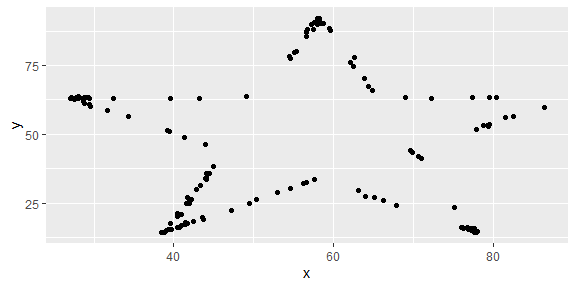
dino\_data %>%  
 summarize(r = cor(x, y))

## # A tibble: 1 x 1  
## r  
## <dbl>  
## 1 -0.0645

### Exercise 3

First let’s plot the data in the star dataset:

star\_data <- datasaurus\_dozen %>%  
 filter(dataset == "star")  
 ggplot(data = star\_data, mapping = aes(x = x, y = y)) +  
 geom\_point()



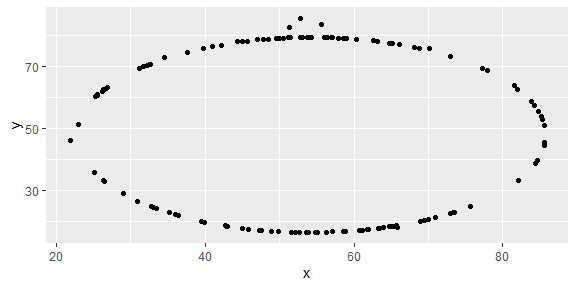
And next calculate the correlation between x and y in this dataset:

star\_data %>%  
 summarize(r = cor(x, y))

## # A tibble: 1 x 1  
## r  
## <dbl>  
## 1 -0.0630

### Exercise 4

circle\_data <- datasaurus\_dozen %>%  
 filter(dataset == "circle")  
 ggplot(data = circle\_data, mapping = aes(x = x, y = y)) +  
 geom\_point()

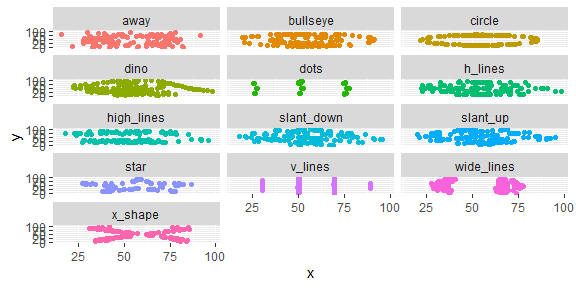


circle\_data %>%  
 summarize(r = cor(x, y))

## # A tibble: 1 x 1  
## r  
## <dbl>  
## 1 -0.0683

### Exercise 5

ggplot(datasaurus\_dozen, aes(x = x, y = y, color = dataset))+  
 geom\_point()+  
 facet\_wrap(~ dataset, ncol = 3) + # facet the dataset variable, in 3 cols  
 theme(legend.position = "none")



datasaurus\_dozen %>%  
 group\_by(dataset) %>%  
 summarize(r = cor(x, y)) %>%  
 print(13)

## # A tibble: 13 x 2  
## dataset r  
## <chr> <dbl>  
## 1 away -0.0641  
## 2 bullseye -0.0686  
## 3 circle -0.0683  
## 4 dino -0.0645  
## 5 dots -0.0603  
## 6 h\_lines -0.0617  
## 7 high\_lines -0.0685  
## 8 slant\_down -0.0690  
## 9 slant\_up -0.0686  
## 10 star -0.0630  
## 11 v\_lines -0.0694  
## 12 wide\_lines -0.0666  
## 13 x\_shape -0.0656