R Project 1 - Hello R

Ashle Markey

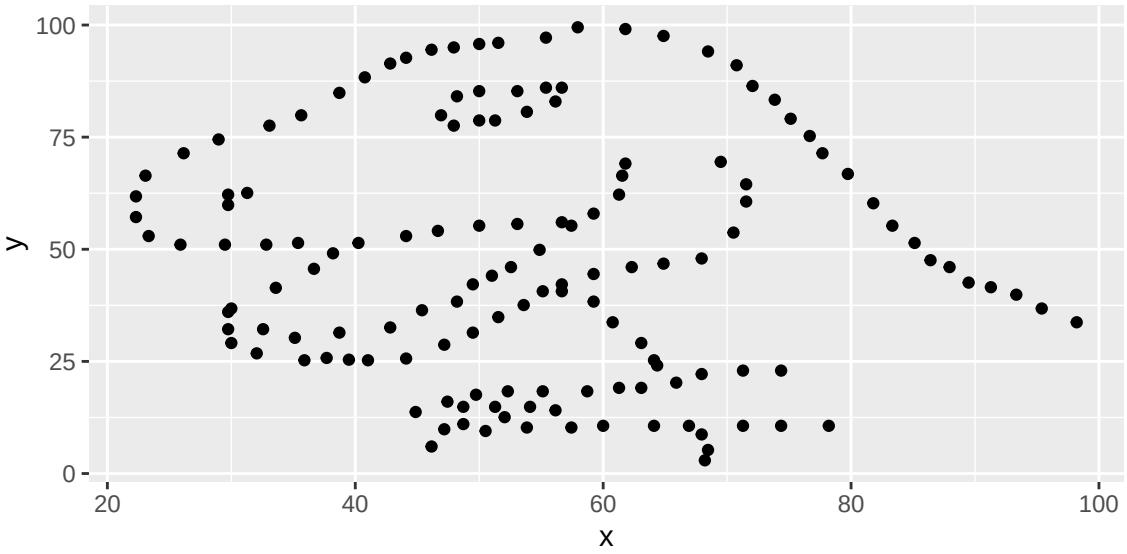
September 8, 2021

Exercise 1

- The datasaurus_dozen file has ***1846 rows*** and ***3 variables (columns)***
- The variables included in the data frame are:
  - **dataset:** indicates which dataset the data are from
  - **x:** x-values
  - **y:** y-values

Exercise 2

First we need to visualize this data to assess the form of the relationship between the variables `x` and `y`. We can do this by plotting the data in the dino dataset:



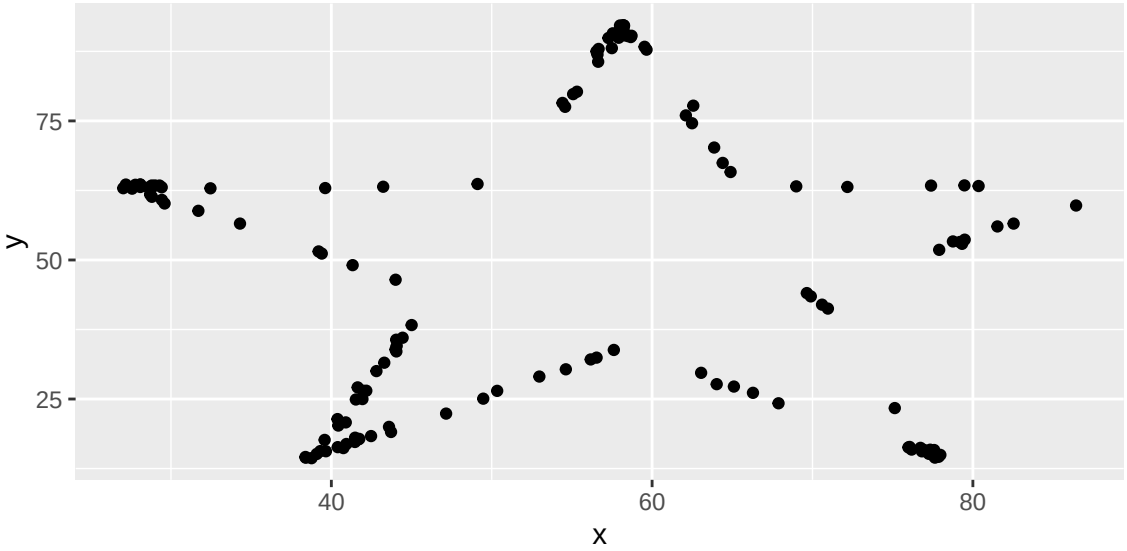
Once we assess the visual relationship between the two variables, we can calculate the correlation between `x` and `y` in this dataset. When we do this, we get:

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

Calculating a correlation coefficient is not particularly needed for this case since the relationship between x and y is not linear.

Exercise 3

We need to visualize and asses the form of the relationship between the variables `x` and `y`for the star dataset. We can do this by plotting the data in the star dataset:



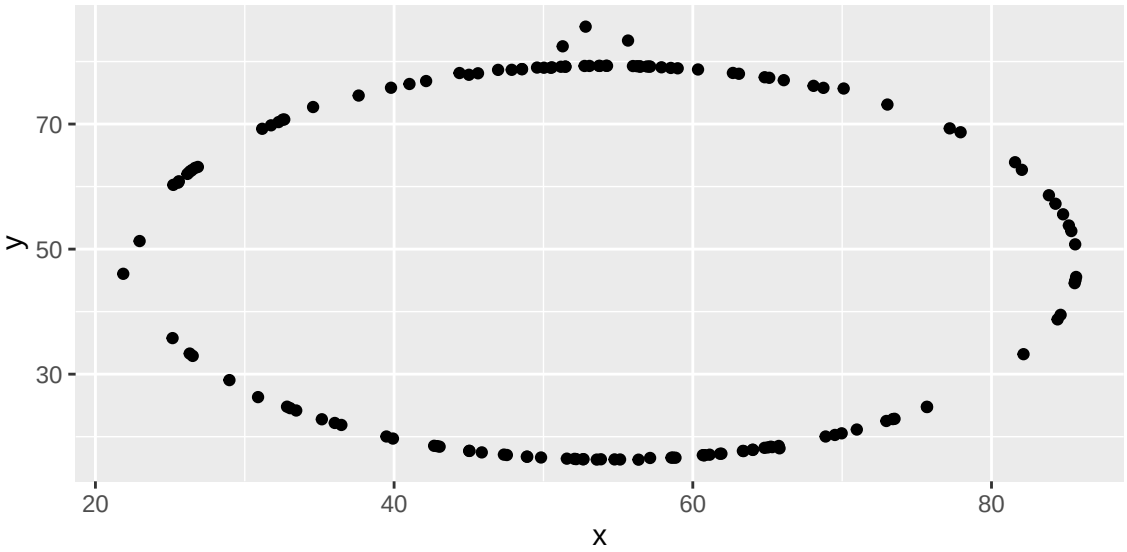
Once we assess the visual relationship between the two variables, we can calculate the correlation between `x` and `y` in this dataset. When we do this, we get:

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

When comparing the value of r in the star dataset to the value of r in the dino data set, you will notice that the r value of star is -0.0630 where as the r value of dino is -0.0645. There is a -0.0015 difference between the two.

Exercise 4

We need to visualize and asses the form of the relationship between the variables `x` and `y`for the circle dataset. We can do this by plotting the data in the circle dataset:



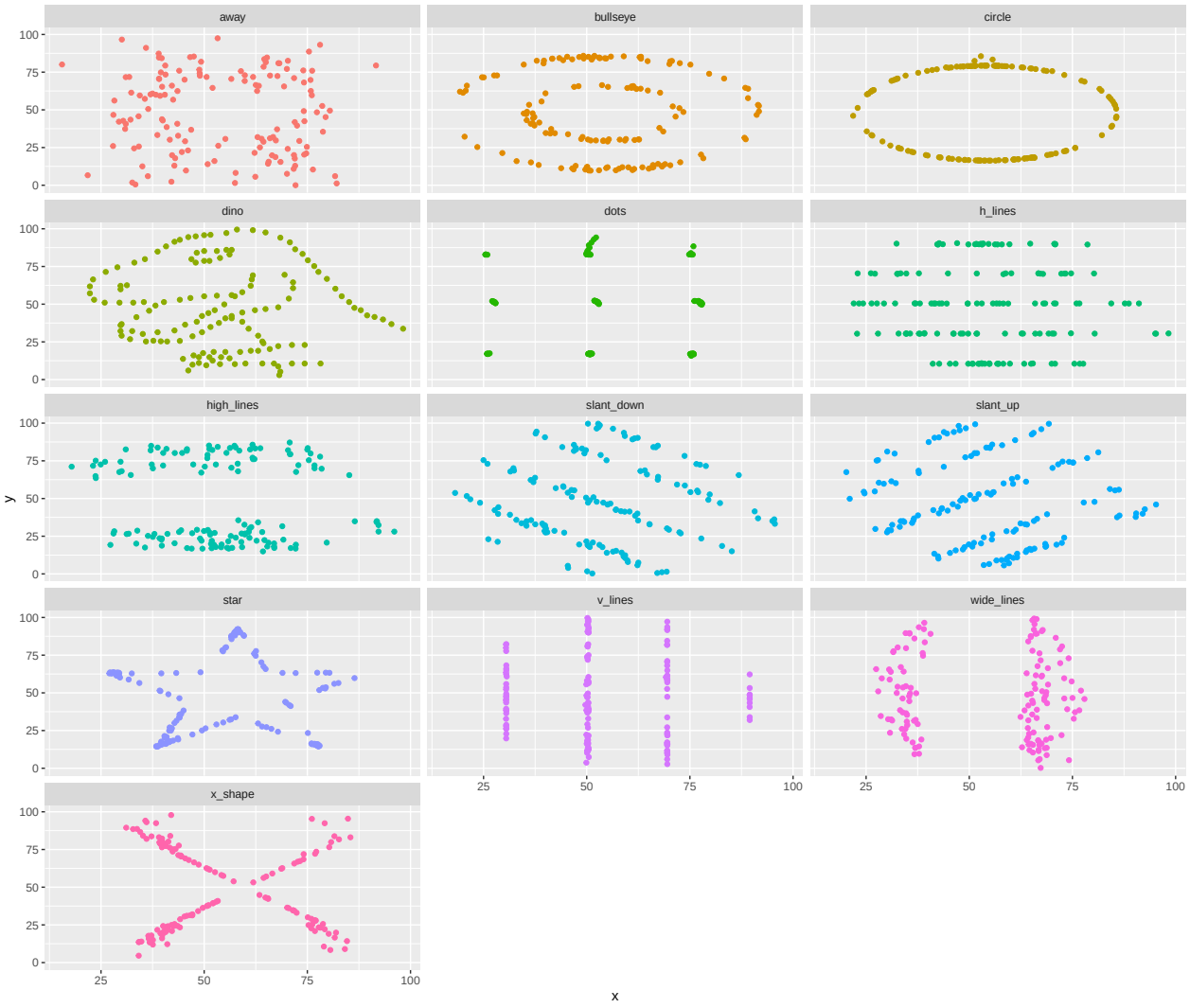
Once we assess the visual relationship between the two variables, we can calculate the correlation between `x` and `y` in this dataset. When we do this, we get:

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

When comparing the value of r in the circle dataset to the value of r in the dino data set, you will notice that the r value of circle is -0.0683 where as the r value of dino is -0.0645. There is a -0.0038 difference between the two.

Exercise 5

We can plot all of the datasets at once and organize them into columns, displaying a more neat visualization for the user.



We can group the summary correlation in a similar manner:

```
## # A tibble:
## #   13 x 2
##    dataset
##    <chr>
```

```
##  1 away
##  2 bullseye
##  3 circle
##  4 dino
##  5 dots
##  6 h_lines
##  7 high_lines
##  8 slant_down
##  9 slant_up
## 10 star
## 11 v_lines
## 12 wide_lines
## 13 x_shape
## # ... with 1
## #   more
## #   variable:
## #   r <dbl>
```