R Project 1 - Hello R

Sharrah Allen-Benson

9/16/2020

Loaded packages:

```
library(tidyverse)
library(datasauRus)
library(magrittr)
library(ggplot2)
library(rlang)
```

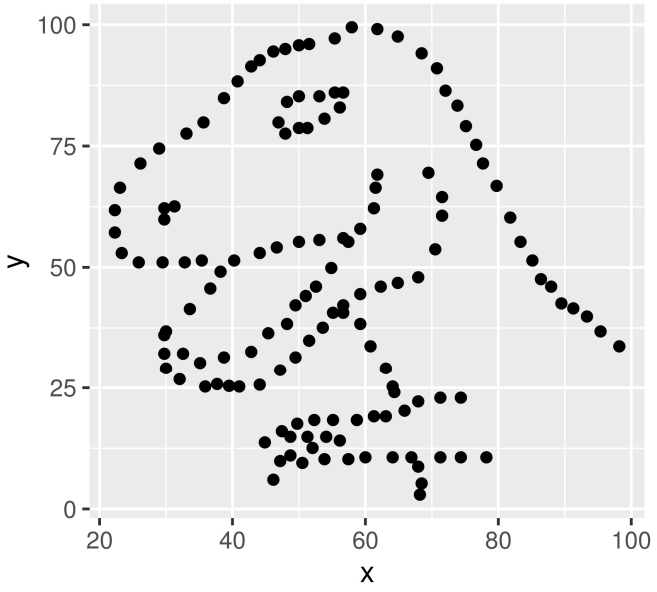
Exercise 1

The datasaurus_dozen file has 1846 rows and 3 columns, or variables. The variables included in the data frame are the x-values, y-values, and dataset.

Exercise 2: dino_data

Here is dino_data plotted:

```
dino_data <- datasaurus_dozen %>%
  filter(dataset == "dino")
ggplot(data = dino_data, mapping = aes(x = x, y = y)) +
  geom_point()
```



Here is the code to calculate the correlation coefficient between x and y for this dataset:

```
dino_data %>%
  summarize(r = cor(x, y))
```

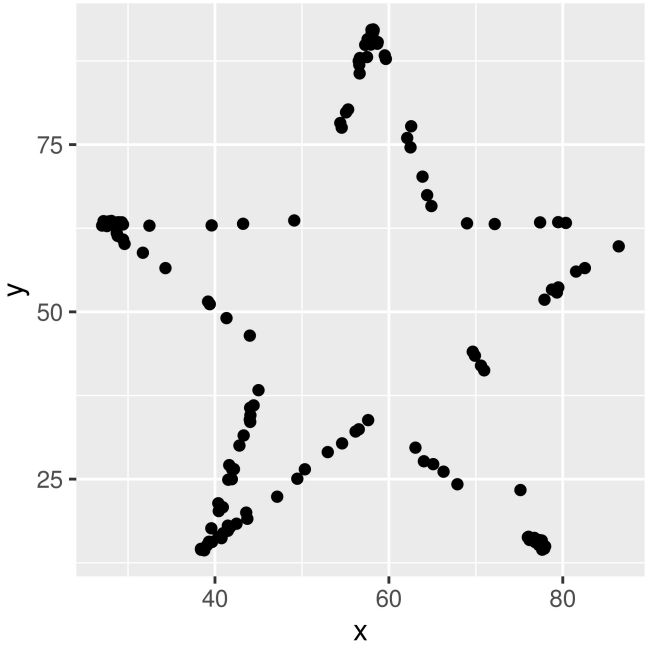
```
## # A tibble: 1 x 1
##         r
##     <dbl>
## 1 -0.0645
```

- The correlation coefficient is -0.0645.

Exercise 3: star_data

Here is star_data plotted:

```
star_data <- datasaurus_dozen %>%
  filter(dataset == "star")
ggplot(data = star_data, mapping = aes(x = x, y = y)) +
  geom_point()
```



Here is the code to calculate the correlation coefficient between x and y for this dataset:

```
star_data %>%
  summarize(r = cor(x, y))
```

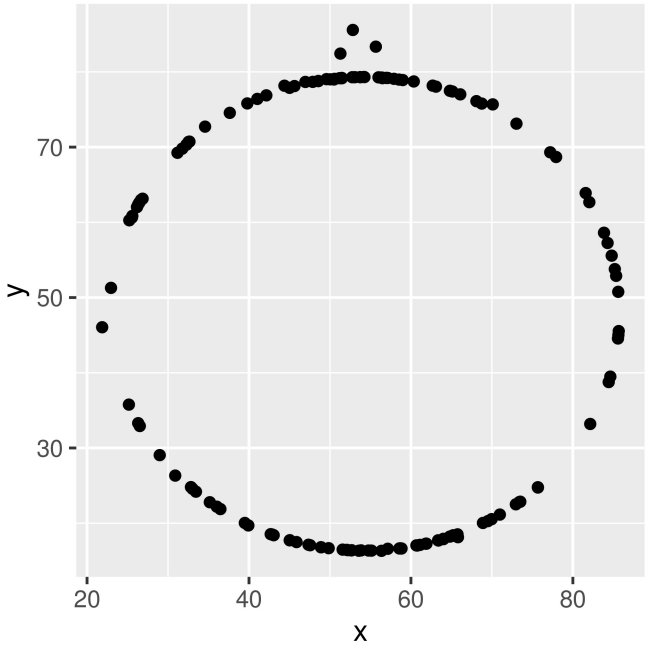
```
## # A tibble: 1 x 1
##         r
##     <dbl>
## 1 -0.0630
```

- The correlation coefficient is -0.0630. Compared to the correlation coefficient of dino_data, -0.0645, this correlation coefficient of -0.0630 is greater than the r of dino_data by 0.0015.

Exercise 4: circle_data

Here is circle_data plotted:

```
circle_data <- datasaurus_dozen %>%
  filter(dataset == "circle")
ggplot(data = circle_data, mapping = aes(x = x, y = y)) +
  geom_point()
```



Here is the code to calculate the correlation coefficient between x and y for this dataset:

```
circle_data %>%
  summarize(r = cor(x, y))
```

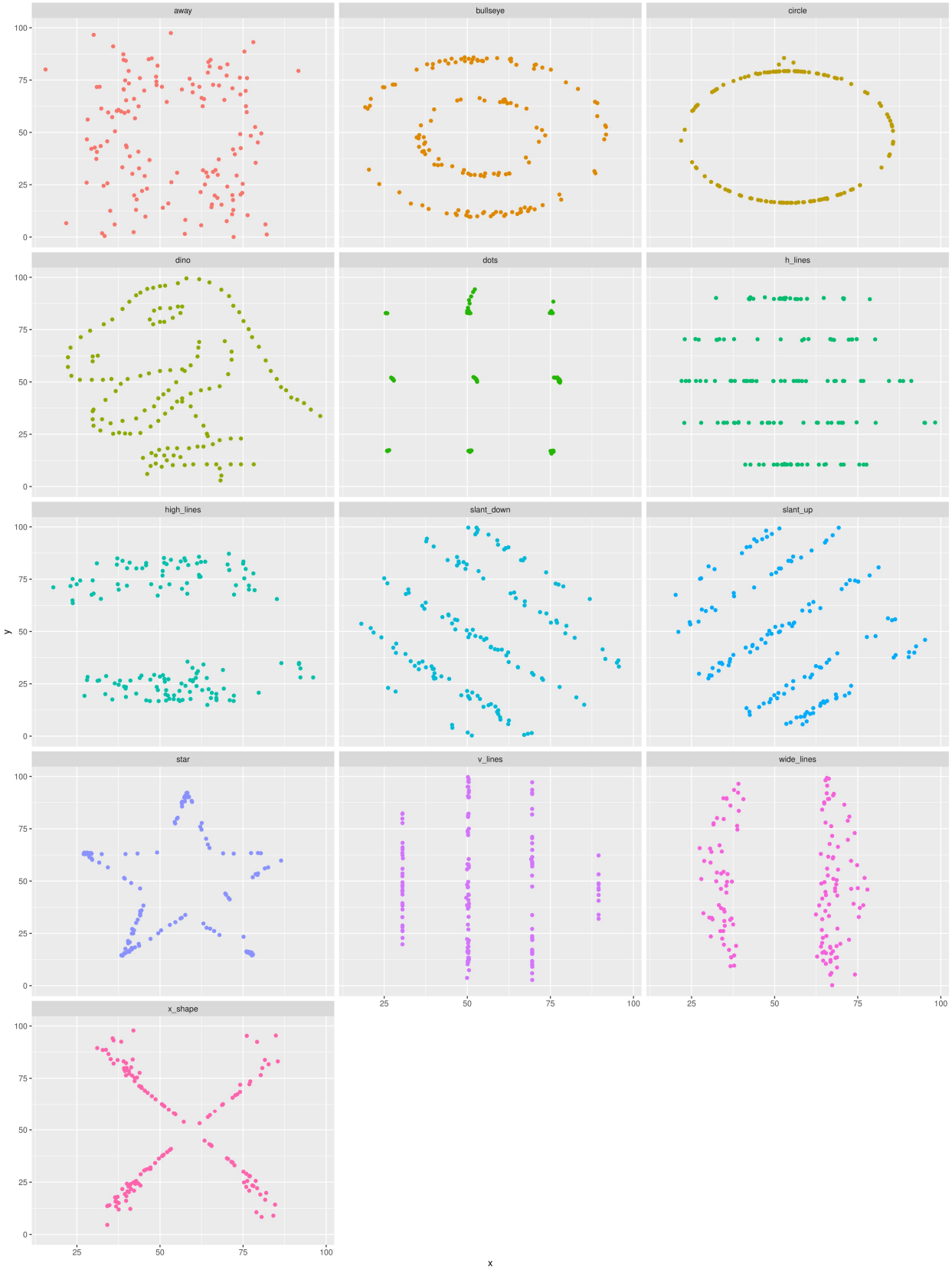
```
## # A tibble: 1 x 1
##         r
##     <dbl>
## 1 -0.0683
```

- The correlation coefficient is -0.0683. Compared to the correlation coefficient of dino_data, -0.0645, this correlation coefficient of -0.0683 is less than the r of dino_data by 0.0038.

Exercise 5: datasaurus_dozen

Here is all datasets in datasaurus_data plotted (plot on next page):

```
ggplot(datasaurus_dozen, aes(x = x, y = y, color = dataset))+
geom_point()+
facet_wrap(~ dataset, ncol = 3) +
theme(legend.position = "none")
```


away
bullseye
circle
dino
dots
h_lines
high_lines
slant_down
slant_up
star
v_lines
wide_lines
x_shape
y
x
0
25
50
75
100

Here is the code to calculate the correlation coefficient between x and y for all datasets in datasaurus_data:

```
datasaurus_dozen %>%
group_by(dataset) %>%
summarize(r = cor(x, y)) %>%
print(13)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # 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
```